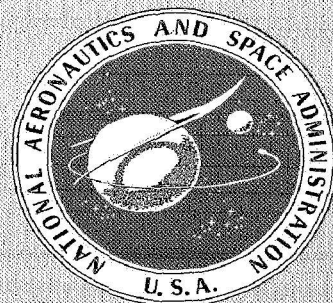


**CASE FILE
COPY**

NASA SP-192

**FOURTH ANNUAL
NASA-UNIVERSITY CONFERENCE ON
MANUAL CONTROL**

A conference held at
UNIVERSITY OF MICHIGAN
Ann Arbor, Michigan
March 21-23, 1968



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

FOURTH ANNUAL NASA-UNIVERSITY CONFERENCE ON MANUAL CONTROL

*University of Michigan
Ann Arbor, Michigan
March 21-23, 1968*



Scientific and Technical Information Division
OFFICE OF TECHNOLOGY UTILIZATION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

1969

FOREWORD

This volume contains the proceedings of the Fourth Annual NASA-University Conference on Manual Control held from March 21-23, 1968, at the University of Michigan, Ann Arbor, Michigan. The program was divided into the following sessions: quasi-linear describing function theory; quasi-linear describing function applications; pilot rating techniques; human performance theory; analysis of adaptive or higher order control processes; system identification, signal analysis, and performance measurement; effects of environmental variables on performance; developments in control and display system design; development of models based on discrete signal processing; and neuromuscular system analysis. Both formal and informal presentations were made; most of the formal and some of the informal papers are included in this volume.

CONTENTS

Page

I. QUASI-LINEAR DESCRIBING FUNCTION THEORY

1.	A Model for Human Controller Remnant	3
	William H. Levison and David L. Kleinman	
2.	The Human as an Optimal Controller and Information Processor	15
	Sheldon Baron and David L. Kleinman	
3.	Mean Square Estimation of Human Pilot Transfer Functions	35
	Richard F. Whitbeck and Frederick D. Newell	
4.	Multimodality Pilot Model for Visual and Motion Cues	47
	R. L. Stapleford	
5.	Synthesis of Human Response in Closed-Loop Tracking Tasks	59
	James J. Adams	

II. QUASI-LINEAR DESCRIBING FUNCTION APPLICATIONS

6.	Application of a Systems Analysis Theory for Manual Control Displays to Aircraft Instrument Landing	69
	W. F. Clement, H. R. Jex, and D. Graham	
7.	An Analysis of Pilot Adaptation in a Simulated Multiloop VTOL Hovering Task	95
	E. W. Vinje and D. P. Miller	
8.	Measurement of Pilot Describing Functions From Flight Test Data with Examples From Gemini X	119
	Rodney C. Wingrove and Frederick G. Edwards	
9.	Models for Steering Control of Motor Vehicles	135
	David H. Weir and Duane T. McRuer	
10.	Man-Machine Models for Car Steering	171
	E. R. F. W. Crossman and H. Szostak	

III. HUMAN PERFORMANCE THEORY

11.	Human Information Processing Rates During Certain Multiaxes Tracking Tasks with a Concurrent Auditory Task	199
	T. E. Wempe and D. L. Baty	
12.	Prediction and Decisionmaking in Manual Control	221
	William R. Ferrell and Harry S. Cohen	
13.	The Effect of Psychological Stress on Decision Processes in a Tracking Task	231
	C. B. Gibbs	

	Page
14. Predictive Compensation in Time-Delay Manual Control Systems	253
Eric A. King-Smith	
15. Psychological and Physiological Skill Development—A Control Engineering Model	275
Ezra S. Krendel and Duane T. McRuer	
16. The Application of Discrete Modeling Elements to the Synthesis and Identification of a Deterministic Model for the Visual Scanning Behavior of Human Operators	289
M. J. Merritt	

IV. ANALYSIS OF ADAPTIVE OR HIGHER ORDER CONTROL PROCESSES

17. Stochastic Modeling of Human Learning Behavior.	315
Jacob L. Meiry	
18. Contributions of Human Behavior as a Multiloop Controller	323
G. Schweizer	
19. Pilot's Response to Stability Augmentation System Failures and Implications for Design	341
David H. Weir and Walter A. Johnson	
20. Model of the Adaptive Behavior of the Human Operator in Response to a Sudden Change in the Control Situation	361
A. V. Phatak and G. A. Bekey	
21. On the Dynamic Response of the Human Operator to Transient Inputs	383
Anil V. Phatak and David H. Weir	
22. State-Space Models of Remote Manipulation Tasks	393
Daniel E. Whitney and Thomas B. Sheridan	

V. SYSTEM IDENTIFICATION, SIGNAL ANALYSIS, AND PERFORMANCE MEASUREMENT

23. Identification of Sampling Intervals in Sampled Data Models of Human Operator.	407
G. A. Bekey and C. B. Neal	
24. Cardinal Reconstruction Theory - A Tool for Estimating Effects of Display Scanning	413
Warren F. Clement	

VI. EFFECTS OF ENVIRONMENTAL VARIABLES ON PERFORMANCE

25. On Biocybernetics of the Vestibular System	433
Laurence R. Young	
26. Motion Cues in Man-Vehicle Control	435
Richard S. Shirley and Laurence R. Young	

VII. DEVELOPMENTS IN CONTROL AND DISPLAY SYSTEM DESIGN

27.	Manual Control System Performance with Predictive Displays	449
	John DeShon Warner	
28.	Some Relations Between Visual and Kinesthetic Displays in Normal Driving.	459
	Malcolm L. Ritchie	
29.	Studies in Driver-Aided Car Following	465
	Robert E. Fenton and Charles L. Sheffer	
30.	An Experimental Steering Simulator for Farm Vehicles	481
	C. Saran and C. W. Suggs	

VIII. DEVELOPMENT OF MODELS BASED ON DISCRETE
SIGNAL PROCESSING

31.	A Revised Stochastic Sampled Data Model for Eye Tracking Movements.	489
	L. R. Young, J. D. Forster, and N. Van Houtte	
32.	Modeling the Human Operator with Finite-State Machines	509
	Lawrence J. Fogel and Roger A. Moore	

IX. NEUROMUSCULAR SYSTEM ANALYSIS

33.	A Closed-Loop Neuromuscular System Explanation of Force Disturbance Regulation and Tremor Data	527
	R. E. Magdaleno and D. T. McRuer	
34.	A Dynamic Model of an Agonist-Antagonist Muscle Pair	543
	W. H. Vickers and T. B. Sheridan	
35.	Effect of External Loading on Human Motor Reflexes	551
	Gyan C. Agarwal, Bradley Berman, Michael T. Hogins, Peter Lohnberg, and Lawrence Stark	
36.	Models of Muscle Proprioceptive Receptors	569
	Gyan Agarwal, Gerald Gottlieb, and Lawrence Stark	

I. QUASI-LINEAR DESCRIBING FUNCTION THEORY

1. A Model for Human Controller Remnant

*William H. Levison and David L. Kleinman
Bolt Beranek and Newman, Inc.*

A model of human controller remnant is developed in this paper and tested against measurements obtained from a variety of manual control experiments. Remnant is defined herein as that portion of the controller's output that is not linearly correlated with the input disturbance to the system. A number of potential sources of remnant are considered and the ability to distinguish between them is discussed. The proposed remnant model is expected to become an integral part of an analysis procedure for predicting the human's visual scanning and control behavior in complex, multivariable control situations.

INTRODUCTION

The remnant is an important component of the quasi-linear representation for the human controller. It frequently accounts for a large fraction of the controller's output, sometimes for most of it (refs. 1 to 3). All too often the remnant is ignored in the analysis of manual control systems. The principal reason for this, we suspect, is the lack of good models for the remnant. We contend that a quantitative understanding of the underlying sources of controller remnant will enhance our ability to predict controller behavior and system performance in multivariable, multidisplay control systems. Two companion papers presented at this conference (refs. 4 and 5) discuss optimal-theoretic models for the controller of multivariable control systems. These models incorporate some of the sources of remnant as inherent constraints on the controller's behavior. The object of this paper is to develop and test a quantitative model of controller remnant based on data obtained from single-variable manual control situations. The work reported herein represents the progress made to date on Contract NAS8-21136, supplied to Bolt Beranek and Newman by NASA-Marshall Space Flight Center.

DEFINITION OF REMNANT

Human controller remnant has generally been defined in the literature as the portion of the controller's response that is not accounted for by his describing function (refs. 1 to 3). The rationale for this definition is based on the assumption that the describing function accounts for all but an insignificant portion of the deterministic component of the control behavior. The remnant then represents the component of the response waveform that is unpredictable except in a statistical sense.

In situations in which the controller's response contains significant nonlinearities or consistent time variations, the above definition of remnant is of questionable relevance. The remnant measured in this situation will represent modeling errors on the part of the experimenter as well as truly randomlike behavior on the part of the controller. A more general approach is to consider remnant as the portion of the controller's response that is not deterministically related to the system inputs. If no constraints on the nature of the deterministic portion of the controller's strategy are implied, the remnant so defined will account solely for the stochastic component of the controller's output.

One practical difficulty that arises from the adoption of the more general definition of remnant is that a very large data base is required for its measurement. Whereas a meaningful describing function can be derived from a single experimental trial a few minutes long, a number of replications of the controller's response to a given input signal are required to separate the deterministic and stochastic components of behavior when no a priori model of the deterministic strategy is assumed. Because of limitations on the manual control data base available to the authors, we shall consider remnant as defined in the conventional quasi-linear model context. We feel that the remnant measurements analyzed in this work are close approximations to the controller remnant as defined in the more general sense, since the experimental conditions that yielded this body of data were designed to minimize nonlinearities and consistent time variations in the controller's strategy (ref. 2).

There are two ways in which to view the remnant in the quasi-linear context. One approach is to consider the remnant as the portion of the controller's response that is not causally related to the perceived system error by the controller's describing function. This concept of remnant is often referred to as the "open-loop remnant," since it is derived from a model of the controller alone. This concept is particularly useful for building theoretical models of controller remnant, but the measurement of open-loop remnant is often difficult when there is a relatively large amount of remnant-induced power circulating around the control loop. Some investigators (refs. 1 and 2) have found that better estimates can be obtained of the "closed-loop remnant," which is defined as the portion of the controller's response not related to the system forcing function. Since the deterministic component of the controller's strategy is considered to be linear and time invariant, the open-loop and closed-loop remnant spectra are related by the linear transformation

$$\Phi_{uu_c}(\omega) = \left| \frac{1}{1+H(\omega)V(\omega)} \right|^2 \Phi_{uu_o}(\omega) \quad (1)$$

where $\Phi_{uu_o}(\omega)$ represents the open-loop remnant spectrum, $\Phi_{uu_c}(\omega)$ the closed-loop spectrum, $H(\omega)$ the controller's describing function, and $V(\omega)$ the vehicle dynamics. Both concepts of remnant are used in this paper.

A MODEL OF CONTROLLER REMNANT

Remnant may arise from a variety of sources such as: (1) errors of observation of the displayed state of the vehicle, (2) errors in executing the intended control movement, (3) time variations in the controller's strategy, and (4) structural deficiencies of the model for the human controller (e.g., deviations of the controller from a continuous, linear, time-invariant control strategy). To simplify the problem, we shall consider in this paper only those control situations that tend to minimize nonlinear and discontinuous control behavior.

We shall, therefore, investigate remnant obtained from manual tracking experiments in which (1) the plant dynamics are linear, (2) the subject is required to view a single display and manipulate a single control, and (3) the task requirements are such that the subject apparently devotes continuous attention to the tracking task. The display may present one or more variables, either linearly correlated or independent, and the subject may obtain additional state variables by performing linear operations on the variables that are explicitly displayed.

The human controller in such a situation can often be represented approximately by the simplified linear model diagrammed in figure 1. The controller's control characteristics are considered as the cascade of three linear operations: a central-processing time delay, a set of adjustable gains, and the neuromuscular dynamics. The time delay is represented for mathematical convenience as occurring entirely in the visual pathways. The adjustable gains represent the means by which the subject attempts to optimize his control strategy to fit a given control situation. The neuromuscular dynamics are considered to be relatively invariant and not a significant source of remnant. Models for all three processes are discussed in reference 1.

Potential sources of remnant are shown in the linear flow diagram of figure 2. The system state vector $\underline{x}(t)$ is delayed by an amount $\tau(t)$ which is assumed to vary randomly about an average delay of τ_0 . To the delayed state vector is added an observation noise vector $\underline{r}_x(t)$. The perceived state $\underline{x}_p(t)$ is processed by the gain vector $\underline{K}(t)$, which varies randomly about an average gain of \underline{K}_0 , and a motor noise component $\underline{r}_u(t)$ is added to yield the "intended" control motion $\underline{u}^*(t)$. The neuromuscular subsystem $H_n(s)$ operates linearly on $\underline{u}^*(t)$ to provide the control motion $\underline{u}(t)$.

The controller's remnant has generally been considered to be noise added to either the controller's input or output (refs. 1 to 3). While this is correct in a mathematical sense, most psychophysiological data indicate that the underlying sources of random behavior are

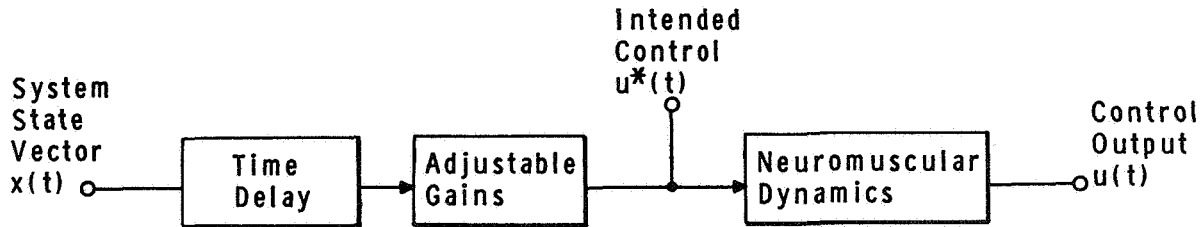


Figure 1.—Simplified model of the human controller in a multivariable, single-control tracking station.

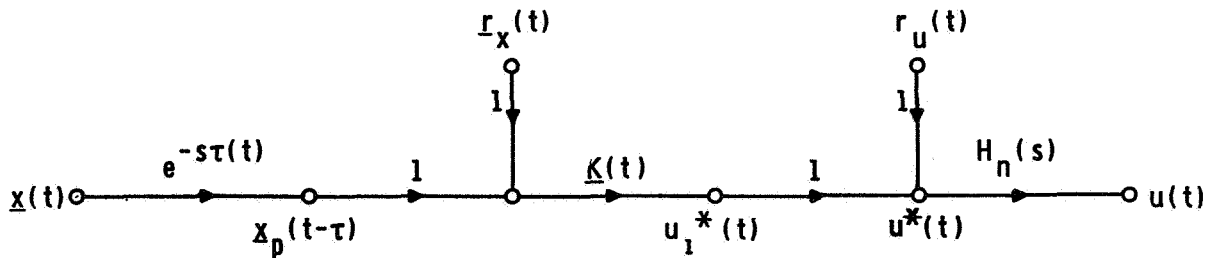


Figure 2.—Linear flow diagram of human controller.

multiplicative, rather than additive (ref. 6). Thus, the injected noise and the time variations are represented as follows:

$$\underline{r}_x(t) = \underline{N}_x(t) \cdot \underline{x}(t - \tau(t)) \quad (2a)$$

$$\underline{r}_u(t) = \underline{n}_u(t) \cdot \underline{u}_1^*(t) \quad (2b)$$

$$\underline{K}(t) = [\underline{I} + \underline{N}_k(t)] \cdot \underline{K}_0 \quad (2c)$$

$$\tau(t) = [1 + n_\tau(t)] \tau_0 \quad (2d)$$

where $\underline{N}_x(t)$ and $\underline{N}_k(t)$ are diagonal noise matrices and $n_u(t)$ and $n_\tau(t)$ are scalar noise terms. These multiplicative noise components are dimensionless quantities, and we assume that they are Gaussian with zero mean and are linearly independent of each other and of the system forcing function.

If the variations of the controller's time delay are small compared to the average delay τ_0 , the first two terms of a Taylor series expansion may be used to obtain the approximate relationship:

$$\underline{x}(t + \Delta T) \doteq \underline{x}(t) + \Delta T \dot{\underline{x}}(t) \quad (3a)$$

Replacing t by $t - \tau_0$ and ΔT by $-n(t)\tau_0$, we obtain

$$\underline{x}(t - \tau(t)) \doteq \underline{x}(t - \tau_0) + n_\tau(t)\tau_0 \dot{\underline{x}}(t - \tau_0) \quad (3b)$$

By applying the relationships of equations (2) and (3) to the solution of the linear flow diagram of figure 2, we compute the following relationship between the state vector and the intended control motion:

$$\begin{aligned} \underline{u}^*(t) = & [1 + n_u(t)] \underline{K}_0' [\underline{I} + \underline{N}_k'(t)] [\underline{I} + \underline{N}_x(t)] \cdot \\ & [\underline{x}(t - \tau_0) + n_\tau(t)\tau_0 \dot{\underline{x}}(t - \tau_0)] \end{aligned} \quad (4)$$

where the symbol (') indicates matrix transposition.

Consideration of only the first-order noise term yields

$$\begin{aligned} \underline{u}^*(t) \doteq & \underline{K}_0' \underline{u}(t - \tau_0) + \underline{K}_0' [n_u(t) + \underline{N}_k'(t) + \underline{N}_x(t)] \underline{x}(t - \tau_0) \\ & - n_\tau(t)\tau_0 \underline{K}_0' \dot{\underline{x}}(t - \tau_0) \end{aligned} \quad (5)$$

For a single-variable control situation, the first-order remnant component is given as

$$\begin{aligned} \delta u^*(t) \doteq & K_0 [n_u(t) + n_k(t) + n_x(t)] x(t - \tau_0) \\ & - n_\tau(t)\tau_0 K \dot{x}(t - \tau_0) \end{aligned} \quad (6)$$

The first term of equation (5) represents the controller's average response characteristics (i.e., the controller's describing function minus the neuromuscular dynamics). The second term indicates the similarity of the effects of observation noise, motor noise, and time variations in controller gains. Observation noise and gain variations are shown to be

indistinguishable, at least in the tracking context under consideration. The motor noise component differs from the observation noise only in its dimensionality (which is unity, since we are considering only single-control situations). It is thus conceivable that the effects of motor noise could be differentiated from those of observation noise by a series of multistate, single-control tracking experiments. Note that motor noise will be indistinguishable from observation noise and gain variations in a manual control situation in which the controller responds primarily to a single state variable—a situation to which most existing manual control data pertain.

Variations in controller time delay, however, appear to contribute to controller's remnant in a somewhat different way than the other processes that we have considered. Whereas the observation noise, motor noise, and gain variation noise terms act as multipliers on the error vector, the time-delay variation term operates on the first derivative of the error vector. Analysis of a series of single-variable, single-control situations should therefore enable us to determine whether time-delay variations or other factors are the dominant sources of remnant.

Since both of the remnant terms of equation (6) contain the scale factor K_0 , we can define for the single-variable control situation an equivalent injected observation noise process that contains the underlying multiplicative sources. We thus define the controller's remnant as

$$\delta u^*(t) = K_0 \cdot r_x^*(t) \quad (7)$$

where the equivalent observation noise is given as

$$r_x^*(t) = [n_u(t) + n_k(t) + n_x(t)] x(t - \tau_0) - n_\tau(t) \tau_0 \dot{x}(t - \tau_0) \quad (8)$$

We have chosen to refer remnant to an equivalent observation noise process, rather than a noise process injected directly at the controller's output, for two reasons. First, there is some indication in the literature (refs. 1 and 3) that the equivalent observation noise is a relatively flat, wide-band process, a property that may facilitate mathematical treatment. Second, the authors have conducted a set of experiments involving peripheral tracking (ref. 2) in which the true observation noise process appears to be an experimental variable.

The notion of an injected observation noise is, at this juncture, strictly a mathematical construct. Clearly, once the controller's describing function has been determined, one can in theory find a waveform which, when injected at the controller's input, will account for the remnant measured in a given experimental trial. The usefulness of this concept depends on whether or not the characteristics of the equivalent observation noise can be predicted from a knowledge of the underlying multiplicative noise sources and of the control environment. The utility of the multiplicative noise concept, in turn, is in proportion to the extent that the noise process represents a basic psychophysiological disturbance process that is invariant with respect to the parameters of the control situation.

In the remainder of this paper we shall discuss techniques for measuring the injected observation noise, present observation noise spectra obtained from manual control data, and attempt to account for these spectra by a model of the underlying multiplicative noise process.

MEASUREMENT TECHNIQUES

To provide a set of datum points against which to test models of controller remnant, the authors have computed observation noise spectra from data obtained from a variety of single-variable, single-control manual tracking experiments. These experiments are described in detail in reference 2. Figure 3 contains a linear flow diagram of the control situation. Vehicle dynamics, input cutoff frequency, and total input power, among other parameters, served as the experimental variables. The vehicle dynamics were either K , K/s , or K/s^2 . The inputs were designed to simulate rectangular noise spectra augmented by a low-power, high-frequency shelf. Input cutoff frequencies investigated were 0.5, 1.0, and 2.0 rad/sec. For most experiments the input power was adjusted to yield a tracking error power of 0.2 cm^2 display deflection. (One centimeter corresponded to about 0.8° deflection of visual arc.)

The input signals were constructed from sinusoidal components to allow us to take advantage of Fourier transform techniques. Use of sinusoidal inputs also facilitated the separation of remnant-induced signals from the linear response to the input, since signal power at other than input frequencies could arise only from controller remnant (except for a small contribution due to imperfect generation of the input and irreducible system noise).

In order to compute observation noise spectra from signals that are directly measurable, we make the following assumptions:

(1) The additive observation noise signal $r_x(t)$ is assumed to be linearly uncorrelated with the input signal $i(t)$.

(2) The remnant-induced power varies continuously with frequency in the vicinity of input frequencies.

(3) Signal power occurring at input frequencies arises almost entirely from the linear portion of the system response and only negligibly from controller remnant.

Because of the way in which we have defined remnant, we are able to compute only the component of $r_x(t)$ that is linearly uncorrelated with the input. Although certain models of the underlying multiplicative sources of remnant predict that $r_x(t)$ will in fact have a component that is correlated with $i(t)$, this component appears to be small compared to the uncorrelated component. McRuer et al. (ref. 1) have shown that the remnant appears to vary smoothly through the input frequencies. The continuity of the remnant spectrum allows us to test the validity of the third assumption in a specific measurement situation. For example, if the control (or error) power measured at a specific input frequency is much greater than

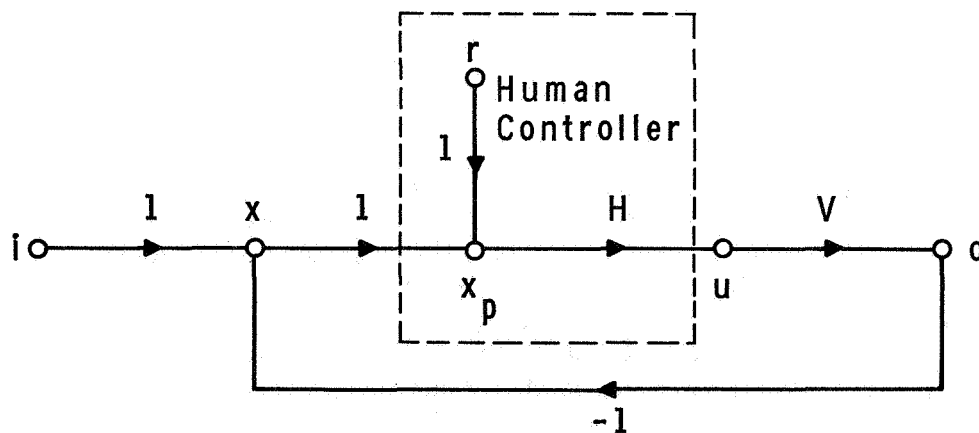


Figure 3.—Linear flow diagram of a single-axis control system.

the remnant-induced power measured at neighboring frequencies, the measurement at the input frequency may be attributed to the input-correlated response of the system. If the remnant-induced power is relatively large, on the other hand, measurements at input frequencies can be expected to include the effects of remnant.

Given that the above assumptions are valid at a particular frequency of interest, the closed-loop error and control spectra may be separated into the following independent input-related and remnant-related components:

$$\Phi_{uu} = \Phi_{uu_i} + \Phi_{uu_r} \quad (9a)$$

$$\Phi_{xx} = \Phi_{xx_i} + \Phi_{xx_r} \quad (9b)$$

where

$$\Phi_{uu_i} = \left| \frac{H}{1 + HV} \right|^2 \Phi_{ii} \quad (9c)$$

$$\Phi_{uu_r} = \left| \frac{H}{1 + HV} \right|^2 \Phi_{rr_x} \quad (9d)$$

$$\Phi_{xx_i} = \left| \frac{1}{1 + HV} \right|^2 \Phi_{ii} \quad (9e)$$

$$\Phi_{xx_r} = \left| \frac{HV}{1 + HV} \right|^2 \Phi_{rr_x} \quad (9f)$$

(The argument ω has been omitted for notational convenience.) Solution of equations (9c) and (9d) yields for the observation noise spectrum:

$$\Phi_{rr_x} = \frac{\Phi_{uu_r}}{\Phi_{uu_i}} \cdot \Phi_{ii} \quad (10)$$

Since measurements of Φ_{ii} and Φ_{uu_i} can be obtained only at input frequencies, the observation noise spectrum obtained in this manner can be specified only at those frequencies. (Although the closed-loop remnant spectrum Φ_{uu_r} cannot be directly measured at input frequencies, a reasonable approximation can be obtained from an average of remnant measurements taken at neighboring frequencies.) At frequencies sufficiently below gain-crossover (i.e., where $|HV| \gg 1$), the observation noise spectrum is approximately identical to the error spectrum at noninput frequencies.

The computation of the scalar observational noise spectrum that is indicated in figure 3 depends neither upon the vehicle dynamics nor upon the controller's describing function and can theoretically be obtained from a single-variable control situation of arbitrary complexity. The interpretation of such a measurement may, however, depend very strongly upon the nature of the control situation.

Consider, for example, a manual control system with K/s plant dynamics. The controller's response will essentially be proportional to a delayed replica of the system error waveform. Since the controller's visual task is therefore to estimate the error waveform, the measured equivalent observation noise may be reasonably interpreted as directly representing, in part, his actual observational errors. When controlling a vehicle with K/s^2 dynamics, on the other hand, the controller responds primarily to vehicle velocity rather than vehicle position. An equivalent observational noise process referred to system error may be less relevant than a noise process referred to error rate. In general, the observation noise process should be treated as a vector process when the vehicle dynamics are more complex than first order.

EQUIVALENT OBSERVATION NOISE SPECTRA

One consequence of the multiplicative noise model of remnant presented in this paper is that the equivalent observation noise scale with the magnitude of the displayed error signal. In order to eliminate differences in spectra due solely to the scaling factor, the observation noise spectra illustrated in this paper are normalized with respect to mean-squared error. Furthermore, the literature supports the notion that normalized spectra will be relatively invariant to the parameters of the tracking situation (refs. 1 to 3).

Figure 4 indicates that the observation noise scales with mean-squared error over a range of almost 10:1. The normalized spectra shown were obtained for mean-squared input levels of 4 and 36 cm^2 display deflection. Since the tracking efficiency was nearly the same for both experiments, the mean-squared system error scaled with mean-squared input. Vehicle dynamics were K/s , and the input cutoff frequency was 0.5 rad/sec. The lack of consistent differences between the two normalized spectra supports the contention that the underlying sources of remnant are basically multiplicative.

Normalized observation noise spectra corresponding to input bandwidths of 0.5, 1.0, and 2.0 rad/sec are compared in figure 5. The vehicle dynamics were K/s , with K the same

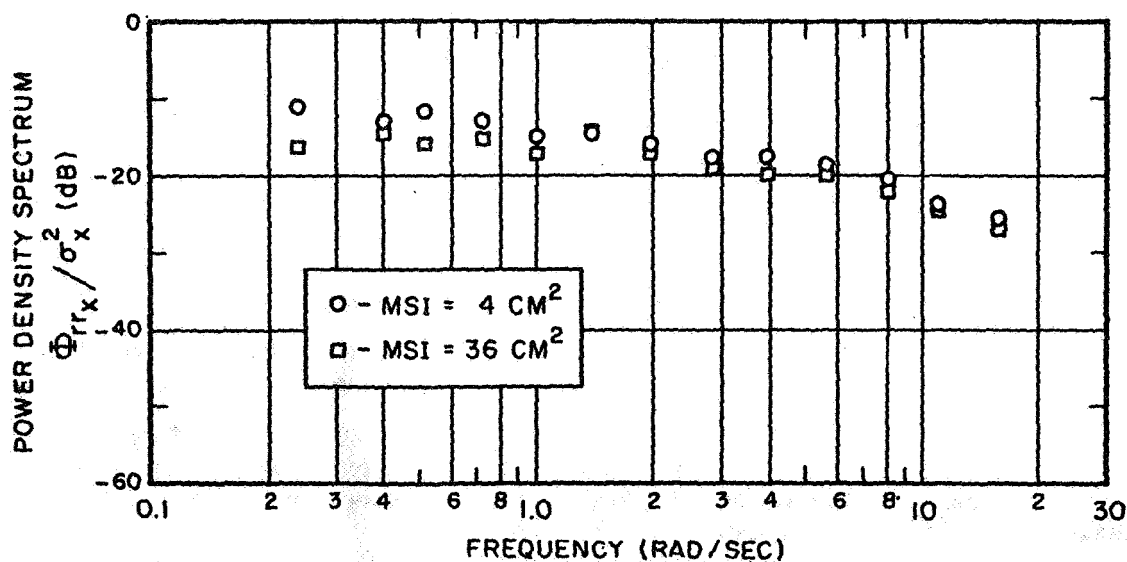


Figure 4.—Effect of mean-squared input on normalized observation of noise spectrum. Dynamics = K/s , Input bandwidth = 0.5 rad/sec. Average of 3 subjects.

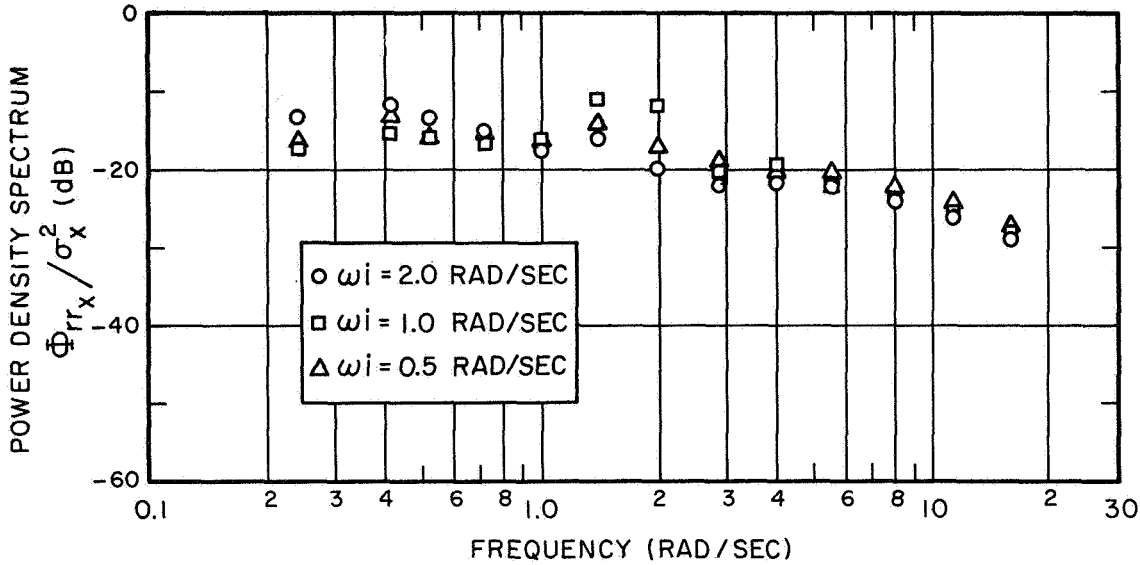


Figure 5.—Effect of input bandwidth on normalized observation noise spectrum.
Dynamics = K/s. Average of 3 subjects.

for all three experiments. The curves differ by less than 4 decibels at all but three frequency points, and what differences there are do not appear to be consistent. These spectra can be described approximately as resembling first-order lowpass noise processes having a cutoff frequency between 4 and 5 rad/sec (i.e., a cutoff frequency approximately equal to the gain-crossover frequency of the tracking loop).

The relative invariance of the normalized observation noise with respect to the parameters of the input disturbance lends further support to the notion that remnant can be attributed to one or more noise processes that are invariant to the parameters of the tracking situation. We have attempted through model-matching techniques to determine whether remnant is due primarily to variations in controller time delay or to one of the remaining processes under consideration: multiplicative observation noise, motor noise, or controller gain variations. Since the latter three processes appear to be indistinguishable in the context of a single-variable control situation, we shall lump them into an equivalent multiplicative observation noise process $n_x^*(t)$ for analytic convenience.

We shall first test the assumption that the equivalent injected observation noise arises primarily from an equivalent multiplicative observation noise process. Equation (8) may then be approximated by

$$r_x^*(t) \doteq n_x^*(t) \cdot x(t - \tau_o) \quad (11)$$

If the linear correlation between $n_x^*(t)$ and $x(t - \tau_o)$ is negligible, the power spectrum of the injected noise $r_x^*(t)$ will be equal to the multiplicative noise spectrum convolved with the spectrum of the error signal.

Theoretical injected noise spectra were computed by the convolution of average error spectra obtained from manual control data with hypothetical lowpass noise spectra of the form

$$\phi_{nn}(\omega) = N_o \frac{\alpha^2}{\alpha^2 + \omega^2} \quad (12)$$

Values of N_0 and α were chosen by the experimenter to achieve the closest match (based on visual comparison) between the theoretical noise spectra and the "measured" spectra presented in figure 5. Theoretical and measured observation noise spectra for input bandwidths of 0.5 and 2.0 rad/sec, respectively, are compared in figures 6 and 7. The following model parameters were used for both computations:

$$N_0 = 0.05 \text{ (rad/sec)}^{-1} \quad (13a)$$

$$\alpha = 1.0 \text{ rad/sec} \quad (13b)$$

A very close match was obtained for the low bandwidth condition; theoretical and measured spectra differed by less than 2 dB at most of the frequency points. The match is less good for the high bandwidth condition, but the theoretical curve appears to reproduce the essentials of the measured observation noise spectrum.

In order to test the assumption that the underlying source of remnant is primarily a variation in the controller's delay, we approximate equation (8) by

$$r_x^*(t) = -n_t(t) \tau_0 \dot{x}(t - \tau_0) \quad (14)$$

The theoretical injected noise spectrum is now given approximately by the convolution of the hypothetical noise spectrum with the spectrum of the error rate, scaled by τ_0 .

Attempts to match the normalized observation noise spectra of figure 5 using the noise model of equation (12) and average error rate spectra obtained from the manual control data have been substantially less successful than the efforts described above, although a wide range of noise parameters has been tried. On the basis of the superior predictions of injected noise spectra obtained from manipulation of error (rather than error-rate) spectra,

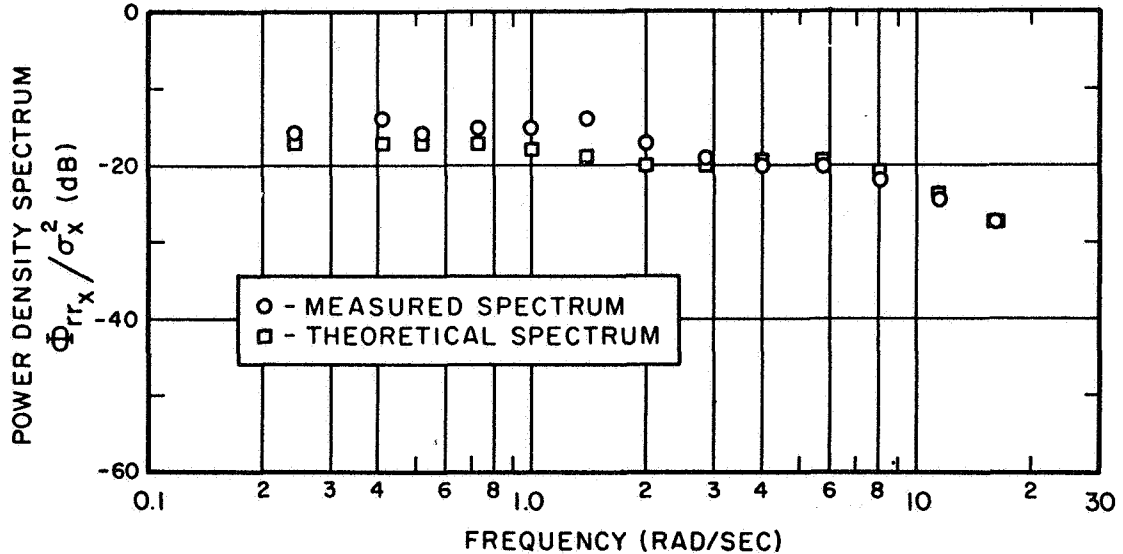


Figure 6. —Measured and theoretical observation noise spectra.

Input bandwidth = 0.5 rad/sec, Vehicle dynamics = K/s.

$$N_0 = 0.05 \text{ (rad/sec)}^{-1}$$

$$\alpha = 1.0 \text{ rad/sec}$$

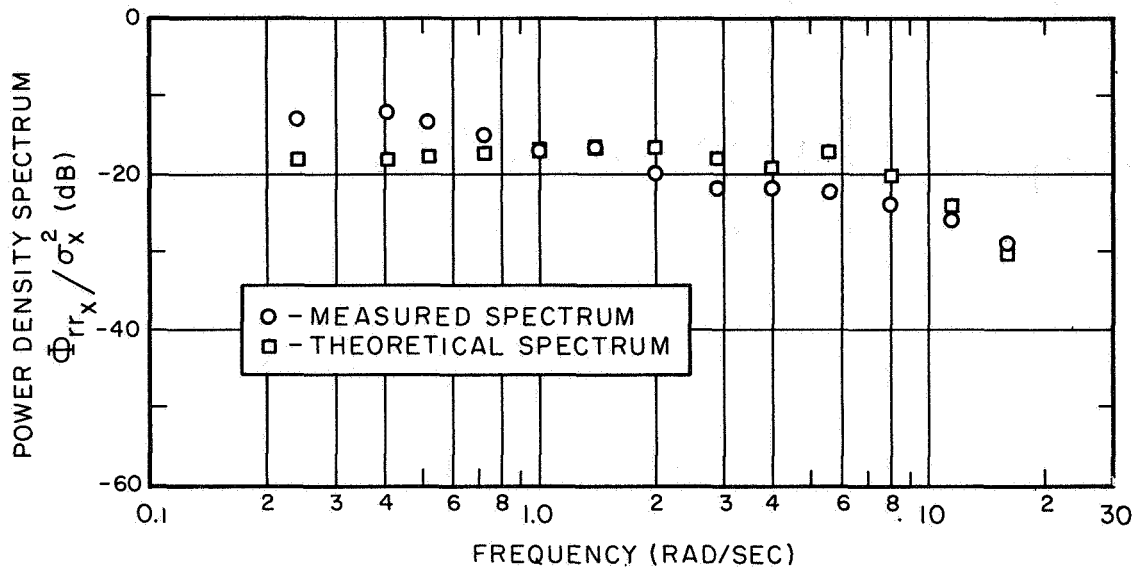


Figure 7. —Measured and theoretical observation noise spectra.
 Input bandwidth = 2.0 rad/sec, Vehicle dynamics = K/s.
 $N_0 = 0.05 \text{ (rad/sec)}^{-1}$
 $\alpha = 1.0 \text{ rad/sec}$

we tentatively assert that time-delay variation is not an important source of remnant. We do not claim to have proven this hypothesis conclusively, since a number of assumptions have been made, and since we have restricted our investigations to a very simple model of the underlying noise process. Nevertheless, the model of an underlying equivalent multiplicative observation noise process seems to be the most parsimonious and may provide the most analytically tractable means of predicting controller remnant in complex control systems. We suspect that further psychophysiological experimentation, not necessarily involving manual control, will be necessary in order to identify and quantify the individual processes contributing to controller remnant.

One additional tracking situation has been investigated to date: single-axis tracking with K/s^2 dynamics and an input bandwidth of 1.0 rad/sec. Since the controller acts primarily as a gain on error-rate, rather than error, in this situation (refs. 1 and 2), we have referred the point of noise injection to error rate. The observation noise spectrum so computed has been normalized with respect to mean-squared error rate and is shown in figure 8 along with a theoretical spectrum. The latter was obtained by convolving the noise spectrum of equation (12) with the average error-rate spectrum. The observation noise spectra obtained for K/s^2 dynamics are wider band than those obtained for K/s dynamics; the parameters of the multiplicative noise spectrum are

$$N_0 = 0.01 \text{ (rad/sec)}^{-1}$$

$$\alpha = 20 \text{ rad/sec}$$

One would not expect the important psychophysiological noise process to be identical in the two situations, since the controller's visual processing tasks differ. Nevertheless, the remarkable similarity of the normalized observation noise spectra suggests a certain

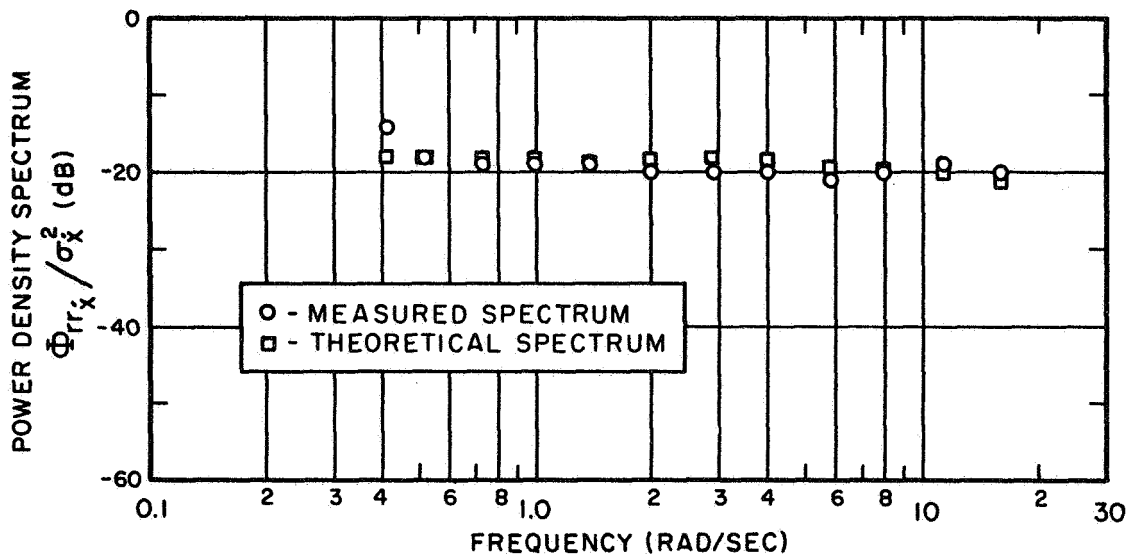


Figure 8.—Measured and theoretical observation noise spectra.

Vehicle dynamics = K/s^2 , Input bandwidth = 1.0 rad/sec.

$N_0 = 0.01 \text{ (rad/sec)}^{-1}$

$\alpha = 20 \text{ rad/sec}$

invariance that may facilitate modeling of human controller remnant in multivariable tracking situations.

ACKNOWLEDGMENTS

The theoretical development and data analysis presented in this paper were supported by NASA-Marshall Space Flight Center under Contract NAS8-21136. The data base upon which the analyses were performed was obtained under Contract NAS2-3080 supplied by NASA-Ames Research Center.

REFERENCES

1. McRuer, D. T.; Graham, D.; Krendel, E. S.; and Reisener, W., Jr.: Human Pilot Dynamics in Compensatory Systems, Theory, Models, and Experiments with Controlled Element and Forcing Function Variations. AFFDL-TR-65-16, July 1965.
2. Levison, W. H.; and Elkind, J. I.: Two-Dimensional Manual Control Systems with Separated Displays. IEEE Transactions on Human Factors in Electronics. Vol. HFE 8, no. 3, Sept. 1967, pp. 202-209.
3. Pew, R. W.; Duffendack, J. C.; and Fensch, L. K.: Summary of Sine-Wave Tracking Studies. Second Annual NASA-University Conference on Manual Control. SP-128, 1966, pp. 15-24.
4. Elkind, J. I.; Falb, P. L.; Kleinman, D. L.; and Levison, Wm. H.: An Optimal Control Method for Predicting Display Requirements and Human Operator Control Characteristics in Complex Vehicular Systems. Paper presented at the Michigan-NASA Conference on Manual Control, March 21-23, 1968.
5. Baron, S.; and Kleinman, D. L.: The Human as an Optimal Controller and Information Processor. Presented at the Michigan NASA Conference on Manual Control, March 21-23, 1968.
6. Green, D. M.; and Swets, J. A.: Signal Detection Theory and Psychophysics. John Wiley & Sons, Inc., 1966.

2. The Human as an Optimal Controller and Information Processor*

*Sheldon Baron and David L. Kleinman
Bolt Beranek and Newman, Inc.*

A mathematical model of the instrument monitoring behavior of the human operator is developed. The model is based on the assumption that the operator behaves as an optimal controller and information processor, subject to his inherent physical limitations. The resulting model depends explicitly on the control task and the control actions. Provision is made for the ability to obtain information from the peripheral visual field. There are no restrictions on signal coupling.

The specific characteristics of the operator's visual sampling behavior are predicted by solving a nonlinear, deterministic optimization problem. A two-axis compensatory tracking example is investigated and the results exhibit the general characteristics expected of a human operator performing a similar task.

INTRODUCTION

We are currently investigating the problem of manual control in complex, multivariable situations with the aim of obtaining a more complete mathematical description of the human operator's control and information processing behavior. The model we are studying is rooted in modern control theory. The main assumption underlying its development is that the well-trained, well-motivated operator behaves in a near optimal manner, subject to certain inherent constraints. The model contains elements for describing the operator's instrument monitoring, data reconstruction and control behavior as well as means for representing his inherent limitations. Thus, the operator's control behavior is assumed to be that of an ideal feedback controller. The human's data-reconstruction process is chosen to obtain a "best" estimate of the state of the controlled element based on information obtained from "sampling" the various instruments.

In this paper we shall be primarily concerned with the instrument monitoring aspects of the human operator model. The approach taken here leads to a representation of the visual sampling process that has some very desirable features not included in previous visual sampling models (ref. 1). Since the visual sampling problem is considered concurrently with the control problem, we find that instrument monitoring behavior depends explicitly on the

*This work was supported by the National Aeronautics and Space Administration under Contract No. NAS12-104.

control task and on the control actions. It is also possible, within this framework, to allow for the ability to obtain information from the peripheral visual field and this ability plays a central role in our development. The development also includes means for explicitly constraining instrument scanning rates.

We began with a brief description of the structure of our model of human operator behavior in a multivariable control task and show how the visual sampling model interfaces with other elements of the overall man-machine system. We find that the specific characteristics of the operators visual sampling behavior are predicted by solving a nonlinear optimization problem. This problem is precisely formulated and methods for its solution are discussed. A simple two-axis compensatory tracking task is then considered and instrument sampling characteristics are obtained. The results exhibit the general characteristics one would expect from a human operator performing a similar task. Finally, we discuss implications of our research efforts and some directions for further work.

THE HUMAN OPERATOR MODEL

The pilot-vehicle-display system considered in this study is shown in figure 1. The pilot's basic task is to control a given vehicle in some prescribed manner. Several outputs may be of concern and the operator has several inputs through which to control the vehicle. It is assumed that the only means available for monitoring the vehicle's performance is a multi-instrument display panel. The salient features of the system of figure 1 are now briefly discussed. A more detailed discussion may be found in reference 1.

VEHICLE DYNAMICS AND DISPLAY.—The vehicle dynamics are represented by the linear, time-invariant equations

$$\dot{\underline{x}}(t) = \underline{A} \underline{x}(t) + \underline{B} \underline{u}(t) + \underline{w}(t) \quad (1)$$

where \underline{x} is the vehicle state vector, \underline{u} the pilot's control or input vector, and where \underline{w} , which represents the external disturbances such as wind gusts, is a zero-mean, gaussian white-noise vector with covariance matrix \underline{W} .¹ We shall assume that the displayed variables $\underline{y}_d(t)$ are linear combinations of the vehicle states. Thus

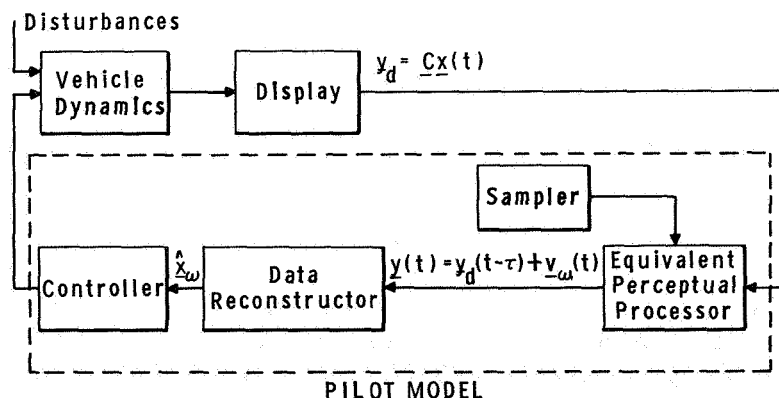


Figure 1.—Model structure.

¹The case when $\underline{w}(t)$ is colored noise can be included in the above formulation by merely introducing additional states.

$$\underline{y}_d(t) = \underline{C} \underline{x}(t). \quad (2)$$

Equation (2) allows for the possibility that all of the vehicles may not be explicitly displayed.

PERCEPTUAL PROCESSOR.—The human operator has various inherent limitations which prevent him from making a perfect instrument reading and from taking instantaneous and precise control actions. These limitations are modeled by combining them into an equivalent perceptual time delay and an equivalent observational noise. The perceptual process is then described by the equation

$$\underline{y}(t) = \underline{y}_d(t-\tau) + \underline{v}_\omega(t) = \underline{C} \underline{x}(t-\tau) + \underline{v}_\omega(t) \quad (3)$$

where $\underline{y}(t)$ is the "observation" or data processed by the human operator, $\underline{y}_d(t)$ are the displayed quantities, τ is the equivalent time delay and $\underline{v}_\omega(t)$ is the equivalent observational noise, and ω represents the human operator's instrument scanning (or sampling) behavior.

We shall assume that \underline{v}_ω is a Gaussian, white-noise process with covariance matrix V_ω that depends on the sampling strategy ω . Numerical values for this matrix will depend on where the pilot is looking (i.e., whether he is viewing an instrument foveally or peripherally) and on relevant features of the display panel. Precise quantitative determination of these values is, of course, an extremely difficult task. However, as we shall see, a course approximation to the values of V_ω may be sufficient in some instances.

CONTROLLER AND DATA RECONSTRUCTOR.—We now give precise meaning to the human operator's control objectives and develop a model for his resulting control and information processing characteristics. We assume that the pilot's control objective is to apply an input $\underline{u}(t)$ to the system (1) so as to minimize a cost functional of the form

$$J(\omega, \underline{u}) = E \left\{ \lim_{t_f \rightarrow \infty} \frac{1}{t_f} \int_0^{t_f} [\underline{x}' \underline{Q} \underline{x} + \underline{u}' \underline{R} \underline{u} + \mathcal{L}(\omega, t)] dt \right\} \quad (4)$$

The constant matrices \underline{Q} and \underline{R} are respectively positive semidefinite and positive definite. The quantity $\mathcal{L}(\omega, t)$ does not explicitly depend on \underline{u} and is included to allow a "direct" cost for sampling. It thus becomes possible to consider, for example, constraints on scanning behavior that arise from human limitations.

Assuming a fixed sampling strategy ω , the human controller's characteristics are therefore determined by the solution of an optimal regulator problem (ref. 2). The solution to this optimization problem is well known (ref. 3) and is characterized by the linear feedback law

$$\underline{u}_\omega^*(t) = -\underline{L}^* \hat{\underline{x}}_\omega(t) \quad (5)$$

The quantity $\hat{\underline{x}}_\omega(t)$ is the least mean-square estimate of the state $\underline{x}(t)$ given the observed data $\underline{y}(s)$, $s \leq t$. It is obtained by reconstructing the system state from the data \underline{y} using a Kalman filter (ref. 4) and a least mean-squared predictor (ref. 5) in cascade (see ref. 1). The feedback gains \underline{L}^* are independent of ω and depend only on the weighting matrices \underline{Q} and \underline{R} and the system dynamics. The minimum cost (for fixed ω) is denoted by

$$J^*(\omega) = \min_{\underline{u}} J(\omega, \underline{u}) = J(\omega, \underline{u}^*) \quad (6)$$

By a sampling strategy we mean a method of deciding which instrument to view directly (foveally) at different time instances. All signals are continuously processed, but, in general, the signal on the sampled instrument will have a lower noise level associated with its reading; this influences the sampling strategy.

The determination of the human operator's sampling strategy will rest on the assumption that the operator behaves in an approximately optimal fashion and samples his instruments accordingly (for example, to minimize estimation error). In other words, the human chooses a sampling strategy ω^* which minimizes $J^*(\omega)$ over the class of admissible strategies, Ω , i.e.,

$$\omega^* = \arg \min_{\omega \in \Omega} J^*(\omega) \quad (7)$$

It is important to note that ω^* will, in general, depend on the control requirements. This will become clear in the next section where a precise formulation of the optimal sampling problem is given.

In order to simplify the mathematical problems associated with determining ω^* in the above manner, it will be assumed that the sampling strategy is independent of the time delay and, therefore, we can solve for ω^* by considering $\tau=0$. Heuristically, this approach seems reasonable. In the situation of pure instrument monitoring, the effect of the time delay upon sampling behavior is minimal at best and a sampling model that ignores this delay can be constructed (as in refs. 9 and 10). Here, in the case of monitoring and control, it should be possible to at least show that the effect of time delay on sampling behavior is of secondary importance and may be neglected. This is currently under investigation.

In retrospect, although the details of the sampling model have yet to be specified, several noteworthy points are already apparent. First, the sampling model is such that the human's monitoring behavior depends upon the control requirements and the control actions in an explicit way. Second, the ability to process information from the peripheral visual field is also included. Finally, there are no apparent restrictions with regard to coupling of signals on various instruments.

OPTIMAL SAMPLING PROBLEM

In the last section we developed the structure for a model of the human operator in a multivariable control and monitoring task. We saw that, given the optimization assumption, we would have to solve an optimal control problem to predict the human operator's characteristics. The optimization problem is twofold in nature. Control is available over not only the plant but also over the manner in which information concerning the system state is obtained and processed. The design of optimal modulating signals for optimal linear estimation, studied by Athans and Schweppe (ref. 6), is a typical representative of this class of problems and has several aspects in common with that of finding optimal sampling strategies. Meier et al. (ref. 7) considered the general problem of controlling a measurement subsystem. (They refer to such problems as measurement-adaptive problems.) They have shown that the optimal choice of a measurement control can be determined, a priori, by solving a nonlinear matrix optimization problem of the type considered earlier by Athans and Schweppe.

The work of Meier et al. is most closely related to our method for determining human sampling behavior. They considered only discrete time problems, but combined optimization and estimation theory (ref. 3) can be used to extend the results to the continuous time, infinite interval case. The results of this extension will be to reduce the sampling problem to a deterministic nonlinear (matrix) optimization problem, the solution of which will provide our

predictions of human visual sampling. Here we discuss only the essential elements of the problem; more detailed information may be found in reference 1.

MATHEMATICAL FORMULATION.—In the ensuing developments we shall characterize sampling behavior as a function of time by making Ω a function space whose elements ω carry the time interval $[0, \infty)$ into a given set $\hat{\Omega}$. The set $\hat{\Omega}$ is the set of values assumed by $\omega(t)$ for $t \in [0, \infty)$,² specifically,

$$\omega(t) \in \hat{\Omega} \quad \text{for all } t \in [0, \infty) \quad (8)$$

In this manner, choosing $\omega(\cdot) \in \Omega$ specifies the sampling behavior as a function of time and, in turn, the time variation of the noise covariance matrix $\underline{V}_\omega = \underline{V}(\omega(t))$.

We recall from section II that the optimal sampling strategy is obtained by minimizing

$$J^*(\omega) = J(\omega, \underline{u}^*)$$

where $\underline{u}^* = -\underline{L}^* \hat{\underline{x}}_\omega$ (see eq. (5)). It is possible to show (ref. 1) for zero time delay, that minimizing $J^*(\omega)$ is equivalent to minimizing

$$I(\omega) = \text{tr}[\underline{L}^* \underline{\Sigma}_{\text{avg}}(\omega) \underline{L}^{*'}] + \mathcal{L}_{\text{avg}} \quad (9)$$

where

$$\underline{\Sigma}_{\text{avg}}(\omega) = \lim_{t_f \rightarrow \infty} \frac{1}{t_f} \int_0^{t_f} \underline{\Sigma}(\omega, t) dt$$

$$\mathcal{L}_{\text{avg}}(\omega) = \lim_{t_f \rightarrow \infty} \frac{1}{t_f} \int_0^{t_f} \mathcal{L}(\omega, t) dt$$

and $\underline{\Sigma}(\omega, t)$ is the covariance matrix of the estimation error

$$\underline{\Sigma}(\omega, t) \triangleq E\{[\underline{x}(t) - \hat{\underline{x}}_\omega(t)][\underline{x}(t) - \hat{\underline{x}}_\omega(t)]'\} \quad (10)$$

and satisfies the equation

$$\frac{d}{dt} \underline{\Sigma}(t) = \underline{A} \underline{\Sigma}(t) + \underline{\Sigma}(t) \underline{A}' + \underline{W} - \underline{\Sigma}(t) \underline{C}' \underline{V}^{-1}(\omega(t)) \underline{C} \underline{\Sigma}(t) \quad (11)$$

with $\underline{\Sigma}(\omega, 0) = 0$.

The cost $I(\omega)$ differs from $J^*(\omega)$ only by a term that is proportional to the driving noise $\underline{w}(t)$, and is independent of the sampling strategy ω . The first term in equation (9) is the cost

²For example, if there are three displays one may choose $\hat{\Omega} = \{1, 2, 3\}$. In this case the value of $\omega(t)$ could correspond to the instrument being viewed foveally at time t . The set Ω consists of piecewise-constant time functions on $[0, \infty)$.

arising from the estimation noise. We note that this term is also a function of the optimal control gains \underline{L}^* . Because these gains are functions of the control task (i.e., the \underline{Q} and \underline{R} weighting matrices in equation (4)), the optimal sampling strategy will depend explicitly on control objectives. The second term in equation (9) represents the sampling cost as discussed earlier.

THE OPTIMAL SAMPLING PROBLEM.—The element $\omega^*(\cdot) \in \Omega$ (or equivalently $\omega^*(t) \in \Omega$ for $t \in [0, \infty)$) that minimizes $I(\omega)$ will be used to predict the pilot's scanning behavior. This is in accordance with our assumption that the pilot acts in a (near) optimal fashion and thereby chooses an optimal sampling pattern.

The mathematical optimization problem of finding ω^* is exceedingly difficult at this stage because of the dependence of $I(\omega)$ upon ω through a time average over $[0, \infty)$. It therefore becomes necessary to introduce further restrictions on the class Ω .

The assumption we shall make is that any element $\omega(\cdot) \in \Omega$ is periodic with period $T < \infty$, where T is fixed but arbitrary. Hence,

$$\omega(t) = \omega(t+T) \quad \text{for all } t \in [0, \infty) \quad (12)$$

The mathematical assumption (12) is tantamount to the physical assumption that the pilot's sampling behavior is periodic with period T .

With this assumption (12) we have, in turn, that $\underline{V}(\omega(t))$ is also periodic with period T . It is then possible to show:

(1) There exists a unique periodic solution of the variance equation (11) having period T . Call this solution $\underline{\Sigma}(\omega, t)$.

(2) If $\underline{\Sigma}(\omega, t)$ denotes the solution of equation (11) with $\underline{\Sigma}(\omega, 0) = \underline{0}$, then $\lim_{t \rightarrow \infty} \underline{\Sigma}(\omega, t) = \underline{\Sigma}(\omega, t)$

The above two statements imply that

$$I(\omega) = \frac{1}{T} \text{tr} \left\{ \underline{L}^* \int_0^T \underline{\Sigma}(\omega, t) dt \underline{L}^{*'} \right\} + \frac{1}{T} \int_0^T \mathcal{L}(\omega, t) dt \quad (13)$$

Consequently, the determination of ω^* has been reduced to the following nonlinear optimization problem:

Optimal Sampling Problem: Given $T < \infty$. Find the element $\omega^* \in \Omega$ which minimizes $I(\omega)$, where $\underline{\Sigma}(\omega, t)$ satisfies

$$\frac{d}{dt} \underline{\Sigma}(t) = \underline{A} \underline{\Sigma}(t) + \underline{\Sigma}(t) \underline{A}' + \underline{W} - \underline{\Sigma}(t) \underline{C}' \underline{V}^{-1}(\omega(t)) \underline{C} \underline{\Sigma}(t) \quad (14)$$

with the boundary condition $\underline{\Sigma}(0) = \underline{\Sigma}(T) > \underline{0}$.

Thus far, the sampling period T has been considered a fixed but arbitrary parameter. However, to completely specify the sampling behavior of the human operator, we must also specify T . This is accomplished, in accordance with our optimization hypothesis, in the following manner. For a particular value of T the optimal sampling problem can, in principle, be solved for ω^* and $I(\omega^*)$. Thus, $\omega^*(\cdot)$ will parametrically depend on T and we write

$$\left. \begin{aligned} \omega_T^*(\cdot) &\equiv \omega^*(\cdot) \\ I(\omega_T^*) &\equiv I(\omega^*) \end{aligned} \right\} \quad (15)$$

We now vary T , and for each T compute $\omega_T^*(\cdot)$ as well as $I(\omega_T^*)$. Thus, $I(\omega_T^*)$ will be a scalar function of T and we let

$$T^* = \arg \min_T I(\omega_T^*) \quad (16)$$

be the requisite choice of scan period (i.e., the optimal scan period). In this manner, ω_T^* will describe the optimal sampling strategy.

To accomplish the minimization indicated in equation (16), it will be necessary to introduce certain nonrepetitive assumptions on the class Ω . As T increases, the set Ω will encompass a wider class of sampling strategies and not merely repetitive cycles of those sampling strategies contained in a set Ω for smaller values of T . (Recall that a function that is periodic in T is also periodic in $2T$.)

It is important to note that the periodicity assumption is not as restrictive as it may first seem. Within a period T , any arbitrary sequence of instrument selections is possible, provided that the entire sequence is not repeated. This means that each instrument may be looked at several times during a single period T and that the interval between these looks is not necessarily the same. The assumption of periodicity is necessitated by the consideration of an infinite time problem and is tantamount to assuming that there will be some steady-state sampling behavior. If we so choose, the optimization framework can be used to predict human sampling behavior for a control task of finite duration and thus obviate the need for the periodicity assumption. This is accomplished by minimizing $J(\omega, \underline{u})$ for a finite t_f (as opposed to taking the limit as $t_f \rightarrow \infty$). The results of such an approach will be very similar to those obtained for the infinite time case, except that t_f replaces T , and the initial condition $\underline{\Sigma}(0)$ must be given in the problem statement. For this case the optimal ω^* will represent the scan pattern for the run duration and not averaged scanning behavior. It should be noted that if t_f is finite the optimal control gains \underline{L}^* are no longer constant.

SOLVING THE OPTIMAL SAMPLING PROBLEM.—The deterministic optimization problem derived above is of a special form. The error co-variance matrix $\underline{\Sigma}(t)$ may be interpreted as a system state, the matrix Riccati equation plays the role of a dynamic system and the function $\omega(t)$ is regarded as a control input. Such a nonlinear optimization problem will normally have to be solved computationally. Several approaches, based on different characterizations of the optimal solution, are possible.

The use of a matrix version of the maximum principle (ref. 8) suggests itself for obtaining necessary conditions for optimality. This was the approach followed by Athans and Schweppe (ref. 6). In addition, there exists the possibility of using dynamic programming methods. These latter techniques might prove quite useful in problems for which the number of instruments to be sampled is not too large.

Although the development of theoretically based computational algorithms for solving the optimal sampling problem is a desirable, if not necessary, goal, it is possible to solve simple versions of the problem by numerical search techniques. The cost $I(\omega_1)$ for a given $\omega_1 \in \Omega$ can always be calculated. It is first necessary to compute $\underline{\Sigma}(0)$ explicitly. $\underline{\Sigma}(\omega, t)$ is

then readily obtained by integrating equation (14). A convergent algorithm for computing $\underline{\Sigma}(0)$, for a given periodic sampling strategy, is given in reference 1. Thus, one can numerically compare the costs $I(\omega_1)$ and $I(\omega_2)$ associated with two different sampling strategies ω_1 and ω_2 . In this manner it becomes possible to perform a numerical search for ω^* for a given value of T . One can then repeat the procedure for various values of T to determine the minimum of I with respect to ω and T . This procedure is used in the next section to determine the optimal sampling strategy for a simple example.

A SIMPLE EXAMPLE

In this section the foregoing concepts and theory are applied to a specific control and monitoring task. In order to illustrate the method and avoid cumbersome numerical calculations, we shall consider a simplified example. We shall illustrate: (1) a method of choosing Ω and the dependence of $\underline{V}(\omega)$ upon ω , (2) a method for constraining sampling rates, and (3) the effects of certain parametric variations upon predicted sampling behavior.

A TWO-AXIS TRACKING PROBLEM.—We assume that a human operator is presented with two separate displays each being driven by the output of a first-order dynamical system of the form

$$\dot{x}_i(t) = a_i x_i(t) + b_i u_i(t) + w_i(t) ; \quad \text{cov}[w_i(t)] = w_{ii} \quad (17)$$

Thus, the two sets of dynamics ($i=1,2$) are completely noninteracting. We assume that x_1 and x_2 are presented on display 1 and 2 respectively, and that the quantities viewed by the human operator will be (cf. equation (3))

$$y_i(t) = x_i(t) + v_i(\omega) ; \quad \text{cov}[v_i(\omega)] = v_{ii}(\omega), \quad i = 1,2 \quad (18)$$

The additive observation noise $v_i(\omega)$ is associated with display i and depends on the sampling behavior $\omega(t)$.³ We defer, for the moment, a discussion of this dependence and of the sampling class Ω .

The control task is to pick $u_1(t)$ and $u_2(t)$ so as to minimize

$$J(u_1, u_2) = \lim_{t_f \rightarrow \infty} E \left\{ \frac{1}{t_f} \int_0^{t_f} [q_1 x_1^2(t) + q_2 x_2^2(t) + u_1^2(t) + u_2^2(t)] dt \right\} \quad (19)$$

Our objective is to predict the human sampling behavior associated with the given control task. We shall accomplish this by solving the optimal sampling problem for ω^* .

CHOICE OF SAMPLING PARAMETERS.—The solution of the optimal sampling problem requires the explicit specification of the class of sampling strategies Ω and the functional relationship between $\underline{V}(\omega)$ and $\omega(t)$. For the example being considered we shall include sampling rate constraints within these specifications.

³It is assumed that $v_1(t)$ and $v_2(t)$ are independent.

Sampling Class.—We consider a fixed (but arbitrary) scan period, T . At any time $t \in [0, T]$ it is assumed that the pilot is either looking foveally at display 1 or display 2, or else, is in the transitional phase (i.e., switching attention). We let $\omega(t)=1, 2$, or 0 to correspond to each of the above cases, respectively. Note that $\omega(t)$ so defined is piecewise constant, and over any time interval may be described by a sequence of values $\{1, 0, 2, 0, 1, \dots\}$. Notice further that sequences of the form $\{\dots, 1, 2, \dots\}$ are not allowed. We assume that every transition of foveal attention between displays requires a finite transition or dead time, $t_0 \approx 0.1$ sec.

We are interested in a sequence in which the pilot samples each instrument once per scan period. Thus, the sequences of values of $\omega(t)$ for any $\omega(\cdot) \in \Omega$ is (assuming for convenience, that instrument 1 is sampled first) $\{1, 0, 2, 0\}$. If we let t_1 and t_2 be the times that a pilot spends in viewing instrument 1 and 2, respectively, we have

$$T = t_1 + t_2 + 2t_0 \quad (20)$$

Thus, any element $\omega \in \Omega$ can be characterized by the single number t_1 (or t_2) since T and t_0 are given. In summary, then, any element $\omega(\cdot) \in \Omega$ is piecewise constant over $[0, T]$ and takes on the sequence of values $\{1, 0, 2, 0\}$.

Characterization of $\underline{V}(\omega)$.—Since $\omega(t)$ can take on only one of three values, $\underline{V}(\omega)$ may be characterized by three matrices $\underline{V}(1)$, $\underline{V}(2)$, and $\underline{V}(0)$. The matrices $\underline{V}(1)$ and $\underline{V}(2)$ correspond, respectively, to the observation noise covariances for foveal viewing of display 1 or display 2, while $\underline{V}(0)$ is the corresponding covariance matrix for switching attention. Since we have assumed independent noise processes, the above matrices are diagonal. Hence, $\underline{V}(\omega) = \text{diag}[v_{11}(\omega), v_{22}(\omega)]$.

Here, we shall assume that the elements of $\underline{V}(1)$ and $\underline{V}(2)$ are given quantities in the problem formulation. These parameters relate to, and are identified with, the display system and its interaction with the human operator. For instance, the numbers $v_{11}(1)$ and $v_{22}(2)$ will be less than $v_{11}(2)$ and $v_{22}(1)$, respectively, since foveal viewing of a given display will result in less observational noise than will peripheral viewing. Similarly, if the display separation is increased (or decreased) we might expect the peripheral noise covariances $v_{11}(2)$ and $v_{22}(1)$ to similarly increase (or decrease).

The elements of $\underline{V}(0)$ will be relatively large since we assume that no display is seen accurately during the transition between instruments. An alternative assumption would be to consider that no instrument is viewed foveally but that some (in this case both) are viewed peripherally during a transition. This latter assumption would also result in relatively large values for the elements of $\underline{V}(0)$. Examination of the variance equation (11) shows that as the elements of \underline{V} increase, the solution, $\underline{\Sigma}$ becomes increasingly independent of the magnitudes of these elements. This implies that a fairly gross approximation to $\underline{V}(0)$ should be adequate for predicting sampling behavior.

OPTIMAL SAMPLING PROBLEM FOR TWO-AXIS TRACKING.—We are now in a position to solve the optimum sampling problem. We are given the system dynamics, cost criterion, class of sampling strategies, and observation noise covariances. It is then easy to show (because display 1 and 2 are independent) that the optimal sampling problem reduces to:

Given $T < \infty$, find the element $\omega^* \in \Omega$ which minimizes

$$I(\omega) = \frac{1}{T} \left\{ x_1^2 \int_0^T \sigma_{11}(t) dt + x_2^2 \int_0^T \sigma_{22}(t) dt \right\} \quad (21)$$

where

$$\left. \begin{aligned} \dot{\sigma}_{11}(t) &= 2a_1\sigma_{11}(t) - w_{11} + \sigma_{11}^2(t)/v_{11}(w(t)) \\ \dot{\sigma}_{22}(t) &= 2a_2\sigma_{22}(t) - w_{22} + \sigma_{22}^2(t)/v_{22}(w(t)) \end{aligned} \right\} \quad (22)$$

$$\sigma_{11}(0) = \sigma_{11}(T) > 0 \quad ; \quad \sigma_{22}(0) = \sigma_{22}(T) > 0 \quad (23)$$

$$l_i = \left(a_i + \sqrt{a_i^2 + q_i b_i^2} \right) / b_i \quad ; \quad i = 1, 2 \quad (24)$$

As discussed above, any $w \in \Omega$ can be characterized by its corresponding value t_1 . Thus, $I(w)$ is a function of the single parameter t_1 and the value of t_1 which minimizes $I(w)$ can be obtained by a simple numerical search. For a fixed value of t_1 the value of $I(w)$ is obtained by first solving equations (22) and (23) for the values of $\sigma_{11}(0)$ and $\sigma_{22}(0)$. This is readily accomplished because $v_{11}(t)$ and $v_{22}(t)$ are merely piecewise constant. We then integrate equations (22) and (21) to obtain $I(w)$.

Thus, for a fixed value of T we can obtain $w^*(\cdot)$. In order to obtain the optimum scan period T it is necessary to investigate the behavior of $I(w^*)$ as a function of T . The value of T which minimizes $I(w^*)$ is denoted by T^* and serves as a prediction of human scan period. By the nature of the class Ω , T^* must necessarily be greater than $2t_0$ (except for the case when foveal attention is directed entirely at one display). Finally, the sampling function w^* which corresponds to T^* will serve to predict mean dwell times for instrument 1 and 2.

For numerical convenience it will be assumed that $v_{11}(0) = v_{11}(2)$ and $v_{22}(0) = v_{22}(1)$, as discussed above.

RESULTS AND DISCUSSION. — We now investigate some of the two-axis sampling behavior predictions of our optimal sampling model. We first examine a "nominal" case and determine the optimal sampling period and the average dwell-times for each instrument. Then we examine the effects of various system changes on the predicted sampling behavior, e.g., (1) changes in controlled element dynamics, (2) changes in noise covariances, and (3) changes in cost functional weightings.

The Nominal Case. — We choose the situation in which both axes are identical, i.e., the dynamics, the driving noise covariances, the associated observation noise matrices, and the cost functional weightings are the same for each axis. The nominal controlled element dynamics correspond to the case $a_1 = a_2 = 0$ (this is equivalent to the k/s dynamics, with $k=1$, frequently used in tracking experiments). The remaining parameters are chosen arbitrarily to be

$$\begin{aligned} b_1 &= b_2 = 1 \quad ; \quad q_1 = q_2 = 10 \\ \underline{V}(1) &= \begin{bmatrix} .05 & 0 \\ 0 & 1.0 \end{bmatrix} \quad ; \quad \underline{V}(2) = \begin{bmatrix} 1.0 & 0 \\ 0 & .05 \end{bmatrix} \quad ; \quad \underline{V}(0) = \begin{bmatrix} 1.0 & 0 \\ 0 & 1.0 \end{bmatrix} \\ \underline{W} &= \begin{bmatrix} .2 & 0 \\ 0 & .2 \end{bmatrix} \end{aligned}$$

A transition time t_0 of 0.1 seconds is assumed throughout.

The results for the nominal case are shown as the circled points in figures 2 to 10. We see that the optimum sampling model predicts, as one would expect, that an equal time is spent on each display. (Note, however, that this is not 50 percent of the sampling period because two-tenths of a second is lost in the switching of attention.) The optimal sampling period T^* is approximately 1.1 seconds. The cost due to sampling $I(w)$ is approximately 2.8.

Effects of Controlled Element Dynamics.—We first examine the effects of varying a_1 and a_2 , that is, the controlled element dynamics, while holding the remaining parameters fixed. The results of such variations are shown in figures 2 to 4. The results agree well with intuition. From figures 2a and 2b we see that a larger percentage of time is spent viewing the axis that is more difficult to track, that is, the axis with the more unstable dynamics. Indeed, as one axis becomes more stable with respect to the other, one soon reaches the point where the best thing to do is to spend all of the time looking foveally at the more unstable axis; the more stable axis is then tracked solely on the basis of information obtained from peripheral viewing of the associated display.

In figure 3 we have plotted the optimal sampling period T^* as a function of a_2 for various fixed values of a_1 . This family of curves has several interesting, and perhaps important, properties. First, for a_1 fixed, the optimal sampling period T^* was found to have a unique minimum when considered as a function of a_2 at $a_2 = a_1$. Moreover, the locus of these minima, i.e., the envelope of the family of curves, seems to be a well-defined, smooth curve. Another interesting point is that T^* has a sharper minimum for values of a_1 corresponding to more stable dynamics. The value of this minimum, however, decreases as the systems become more unstable. The decreasing minimum sampling period can be explained by noting that as both systems become more and more unstable it becomes necessary to view each display foveally more often to avoid large estimation errors (associated with peripheral

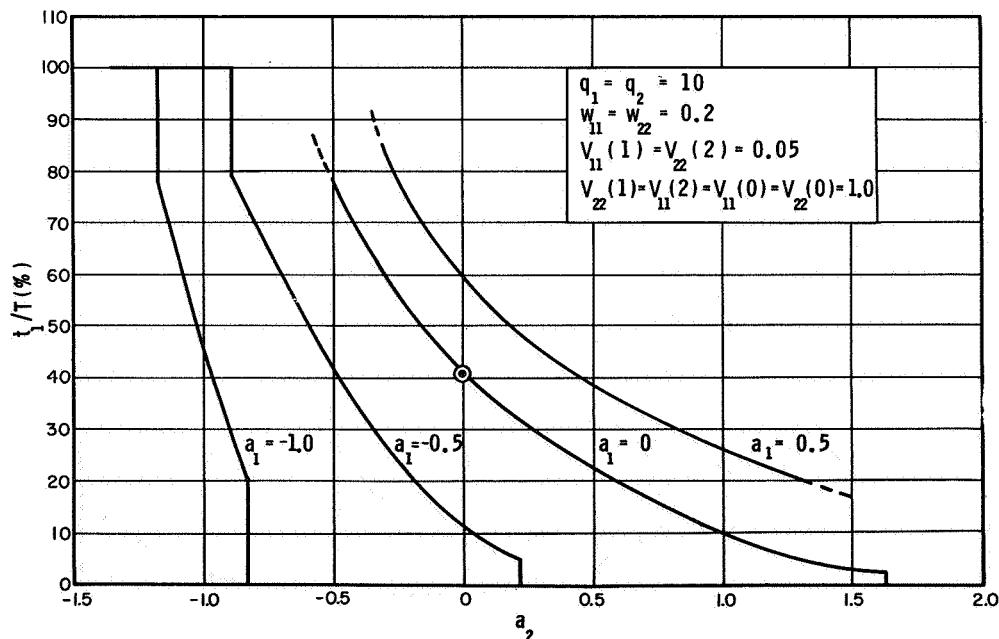


Figure 2a. —Effect of controlled element dynamics on fractional time spent on instrument 1.

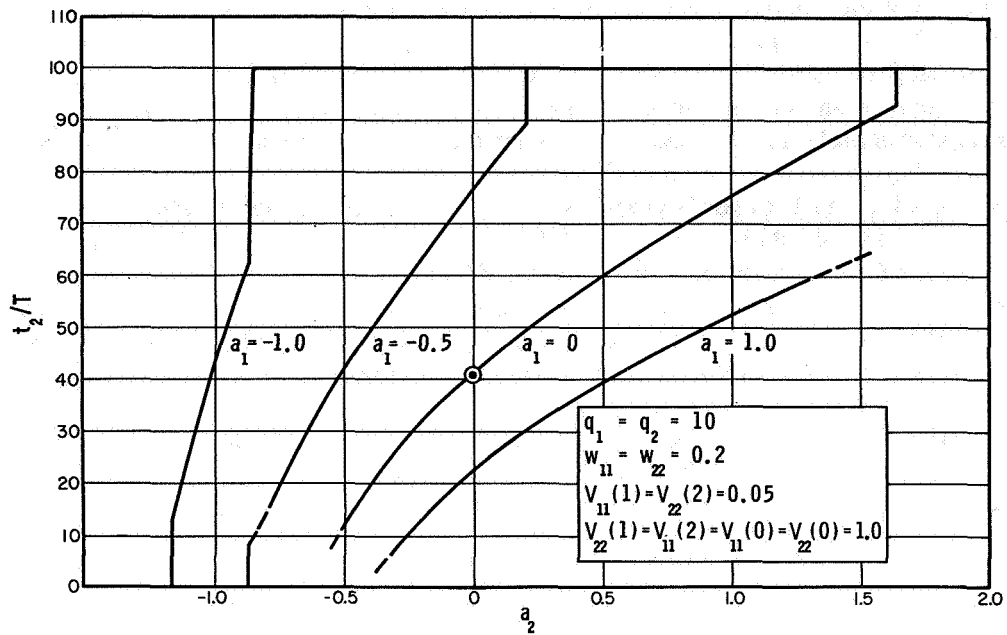


Figure 2b. —Effect of controlled element dynamics on fractional time spent on instrument 2.

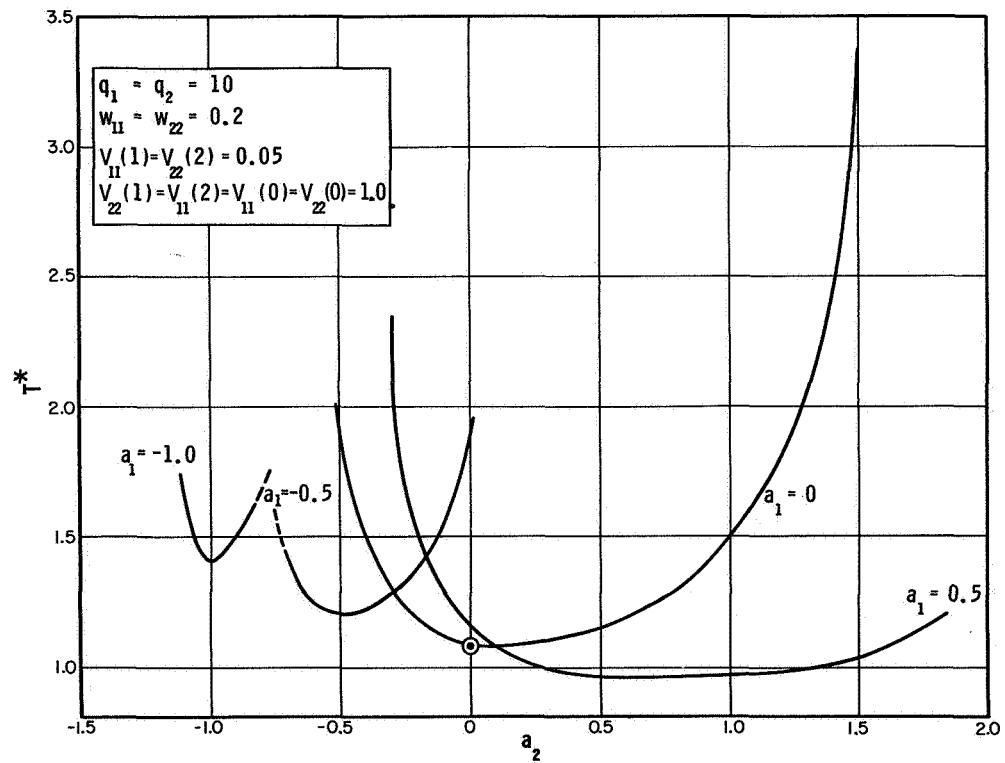


Figure 3. —Effect of controlled element dynamics on optimal sampling period.

viewing).⁴ The sharpness of the minima is more difficult to explain. It may be related to the decreasing value of the minimal optimal sampling period; thus, as $\text{Min } T^*$ decreases, that is, as the systems become more unstable, the relative contribution of the dead time $2t_0$ to the sampling period increases and, consequently, small variations in the dynamics of one system do not appreciably alter the optimal sampling period.

Finally, the effects of changes in controlled element dynamics on the cost for sampling is illustrated in figure 4. As expected the costs increase as the systems become more unstable.

Effects of Observational Noise Covariances.—As noted earlier, the observational noise is used to model the effects of imprecise measurements resulting from display related phenomena or inherent pilot limitations. Here, we examine changes in sampling behavior that result from variations in both foveal and peripheral noise levels, as well as in the ratio of the two. All other system parameters are fixed at their nominal levels for this investigation.

Figures 5 and 6 show the effects of changes in peripheral and foveal noise levels on the optimal sampling period and the sampling cost, respectively. Several conclusions may be drawn from examination of these figures. First, for fixed values of peripheral noise, increases in the foveal noise result in increases in the sampling period. This may be explained by the fact that as the foveal noise level approaches that of the peripheral noise, the quality of information obtained by foveal viewing degrades to the point where it is not much better than that which can be obtained by peripheral viewing. Consequently, there is less and less

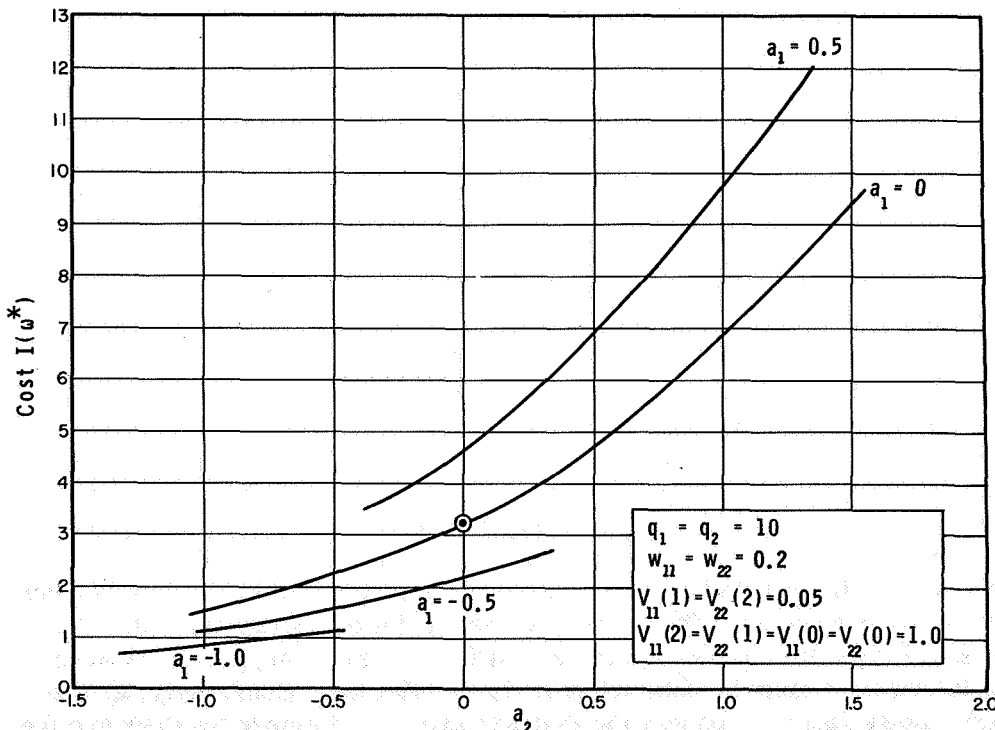


Figure 4.—Effect of controlled element dynamics on minimum sampling costs.

⁴Recall that scanning frequency is inversely proportional to T^* .

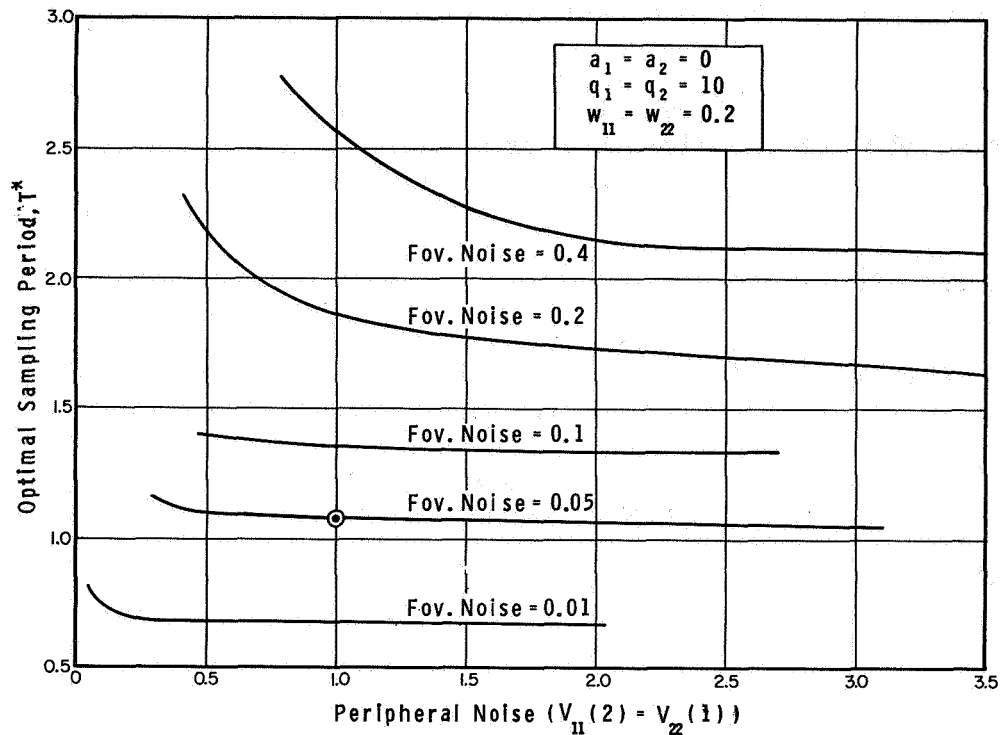


Figure 5.—Effect of observation noise on optimal sampling period.

reason to switch foveal attention. In the limit, when the two noise levels are equal, there is no reason to switch foveal attention and one would expect the sampling period to approach infinity. The curves of figure 5 do indeed show very rapid increases in the sampling period as the foveal noise level approaches the peripheral noise level. On the other hand, for fixed foveal noise levels we find that as the peripheral noise increases, the sampling period decreases and rapidly tends to some nearly constant value. The fact that slight changes in large peripheral noise levels do not appreciably affect the sampling period is not surprising and was discussed earlier. As peripheral noise levels increase the amount of information that can be obtained from peripheral viewing decreases. Beyond a certain point, the only useful information is obtained during foveal viewing and further increases in peripheral noise levels have negligible effect on sampling behavior.

An appealing interpretation of the above results is possible. The foveal noise level can be considered to be a function of the display design. High foveal noise levels correspond to a poor display. Thus, if the foveal noise is high, longer reading times and, consequently, longer sampling periods are required. The ratio of foveal to peripheral noise, for a fixed foveal noise level, can be interpreted as a measure of display separation. As the displays are separated this ratio decreases. Thus, one reaches a point of display separation beyond which increases in the separation do not appreciably affect the sampling period because nearly all useful information must be obtained from foveal viewing. Conversely, as the foveal to peripheral noise ratio increases, the displays are moved closer together and the necessity for visual sampling decreases. When the ratio is one, the instruments may be considered superimposed or integrated. Then, there is no necessity for visual sampling and $T^* \rightarrow \infty$.

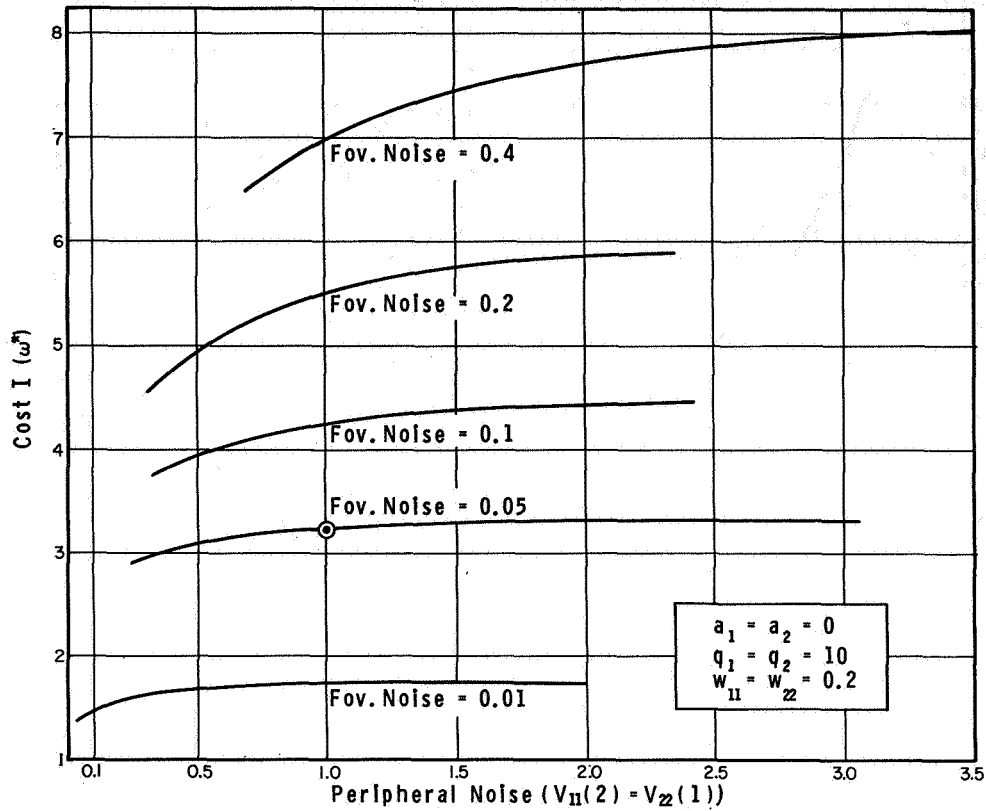


Figure 6.—Effect of observation noise on sampling costs.

Finally, the cost $I(\omega)$ increases monotonically with increases in either foveal or peripheral noise. The costs also tend to level off as the peripheral noise increases so that the sensitivity of the cost to changes in peripheral noise level, for large peripheral noise levels, is small. These results are consistent with our expectations.

Effects of Input Noise Covariance.—For fixed observational noise, the effects caused by changes in input noise levels are illustrated in figures 7 and 8. Figure 7 shows that as the input noise level increases, the sampling period T^* decreases. Increasing the input noise covariance may be interpreted as making the task more difficult. This, as in other instances, results in more frequent sampling. As expected, the effect is similar to that obtained when the input noise level is fixed and the observation noise is decreased. It would appear that the significant parameter is the ratio of input and observation noise covariances.

In figure 8, we see the effects of having different input noise levels on each axis. In particular, we have plotted the percent time spent on instrument 2 versus the driving noise level on axis 2 (w_{22}) for various values of axis 1 input noise levels (w_{11}). We find that for fixed values of w_{11} the percent time spent on instrument 2 increases as w_{22} increases. This apparently results from the fact that the control task becomes more difficult as the input noise level increases.

Effects of Control Requirements.—One of the more significant aspects of the optimal sampling model is the interaction between control requirements and sampling behavior. In the optimization framework we are using, control requirements are explicitly stated in terms of a selection for the weighting matrices \underline{Q} and \underline{R} . We have normalized this selection by letting \underline{R} be the identity matrix. Thus, for the two axis tracking problem we need

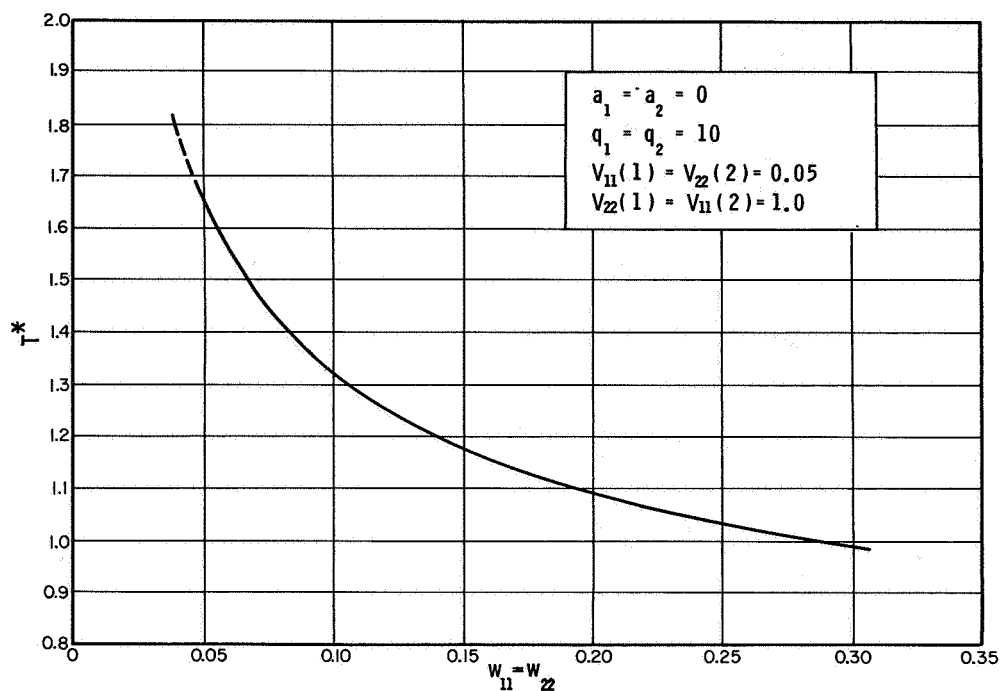


Figure 7. —Effect of input noise covariance on sampling period.

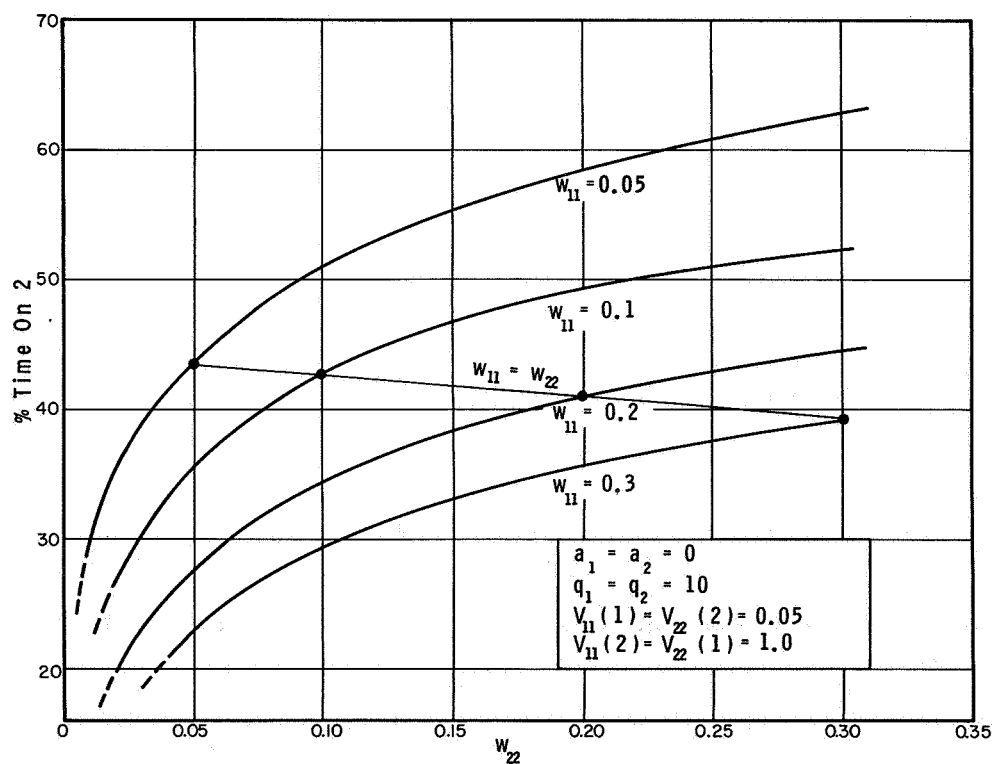


Figure 8. —Effect of input noise covariance on fractional allocation of attention.

only choose q_1 and q_2 (since Q is assumed diagonal). The most significant quantity is actually the ratio of these two weighting coefficients for this simple case.

Figures 9 and 10 show the effects of changing the ratio q_2/q_1 , for fixed q_1 , on the fraction of time spent on instrument 2 and the optimal sampling period, respectively. The results are as expected. Figure 9 shows that as q_2 decreases, i.e., as control of x_2 becomes less important, the time spent on instrument 2 also decreases, in spite of the fact that the two axes are identical in all other respects. It is interesting to note that the point at which instrument 2 is not looked at, at all, occurs prior to $q_2=0$. This is apparently due to the inclusion of a cost for switching attention ($2t_0$). Thus, as q_2 decreases, a point is reached where the penalty for switching attention exceeds the benefits of obtaining better information, by foveal viewing, concerning x_2 .

From figure 10, we see that the optimal sampling period T^* is a minimum when $q_2/q_1=1$, i.e., when both states are equally weighted. In addition, small departures from equal weighting do not materially increase the sampling period. As q_2/q_1 becomes either very small or very large the sampling period increases rapidly. This is to be expected because both cases correspond to a situation in which one instrument is not very important and, hence, there is little need to look at it.

CONCLUSIONS

A model of the human operator in a multivariable control and monitoring task has been developed by considering the human as an information processing-control system. The key idea underlying the model is the assumption that the well-trained, well-motivated human

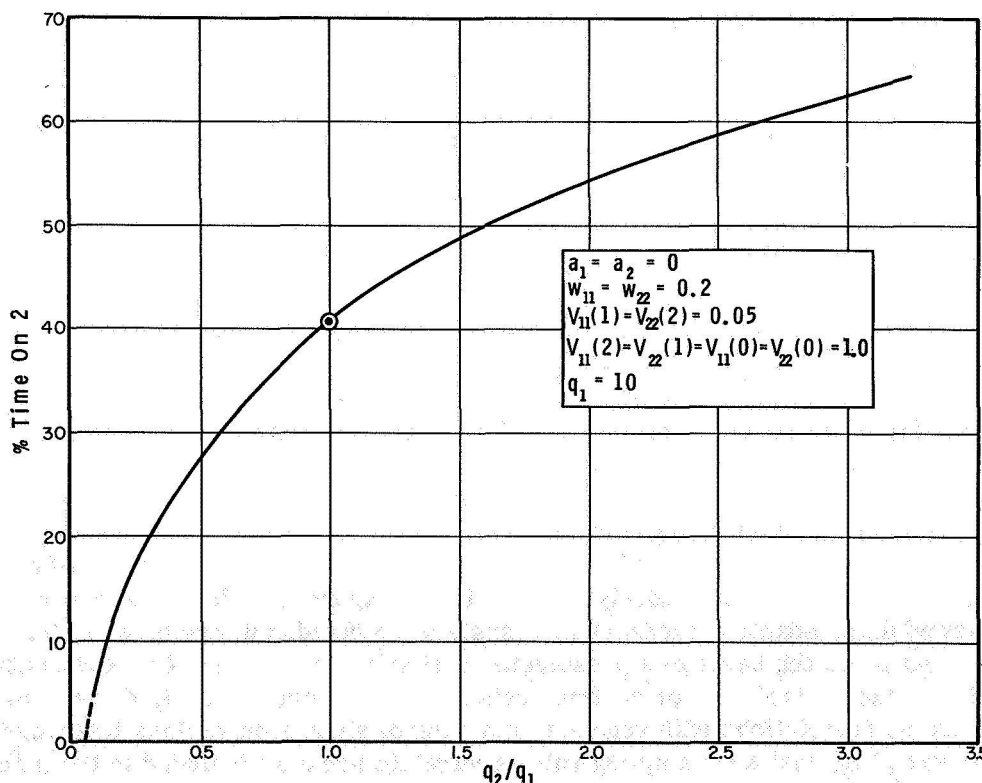


Figure 9.—Effect of control requirements on fractional time spent on instrument 1.

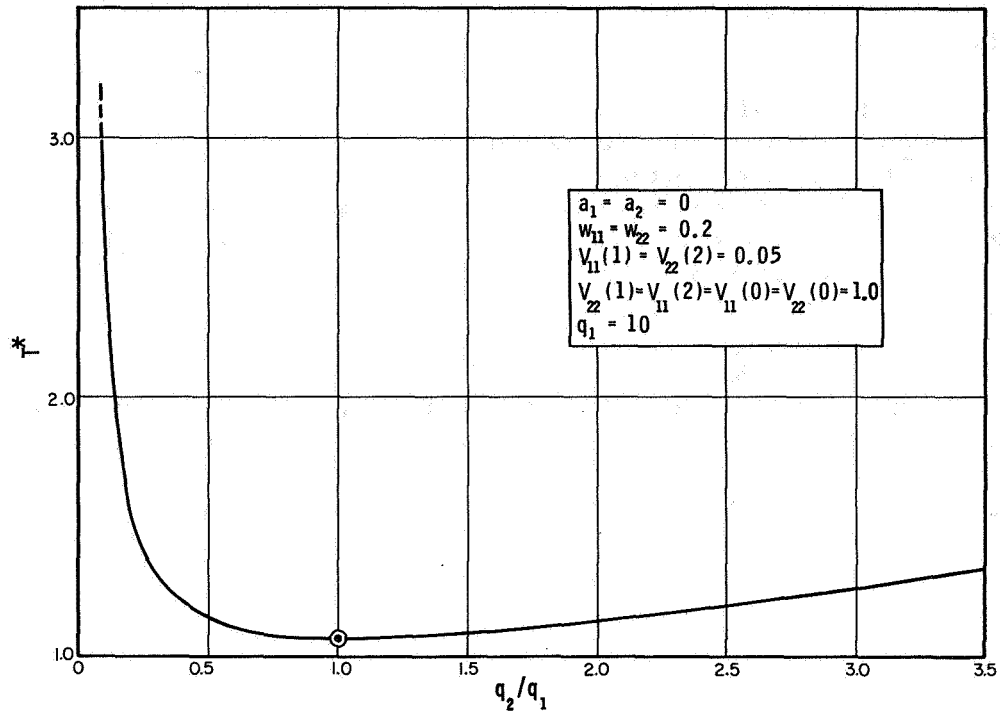


Figure 10. —Effect of control requirements on sampling period.

operator behaves in a near optimal manner, subject to his inherent limitations. The model contains an element for describing the human's limitations in addition to representations of his instrument monitoring, data reconstruction, and control behavior.

The human's limitations are modeled by combining them into an equivalent perceptual time delay and an equivalent observational noise. The observation noise covariance depends on where the operator looks and, hence, allows the effects of peripheral vision on tracking and instrument sampling behavior to be modeled.

The operator's control behavior is assumed to be similar to that of an ideal feedback controller. The feedback gains, which multiply estimates of the system states, are selected so as to minimize a weighted sum of mean-squared tracking errors and control effort.

It is assumed that the human operator processes the information he obtains from scanning the display panel so as to reconstruct a best estimate of the data necessary for control. The data reconstructor consists of a Kalman estimator and a least mean-squared predictor in tandem. The structure of the data reconstructor is fixed but its parameters depend on the operator's instrument sampling strategy.

We postulated a new sampling model that was consistent with other portions of our model in that it was based upon an underlying optimization approach. In this approach the operator is viewed as an adaptive-measurement system; he decides where he will direct his foveal visual attention on the basis of a preselected optimality criterion. This criterion depends explicitly on the control task or control requirements. Peripheral vision is accounted for and there are no restrictions with regard to coupling of signals on various instruments. Provisions for sampling cost and sampling rate constraints are also included in the model.

The assumption of the optimality of the human operator leads to the necessity for solving a multifaceted optimization problem to predict specific operator characteristics. This problem has been formulated precisely and methods for obtaining its solution have been discussed. Finally, we examined a simple two-axis compensatory tracking problem and used the model to predict instrument sampling behavior. The results exhibit the general characteristics one would expect from a human operator performing a similar task.

SOME TOPICS FOR FURTHER INVESTIGATION.—Thus far, we have postulated a model for the human operator in a controlled environment that has interesting and intuitively appealing properties. We have not, however, verified that this model is a valid representation of human behavior. Previous experiments (ref. 5) have shown that the controller and predictor portions of the model yield reasonable representations of the human operator. It is now necessary to determine the efficacy of the sampling and estimation models. Hence, the next step in our research program will be to test these models against existing experimental data.

In the present study the observational noise covariances were chosen arbitrarily. It is our hope that numerical values for these quantities can be measured in appropriate experiments. Because the noise covariances depend on the display and the environment as well as on intrinsic human properties, their determination will be extremely difficult.

In addition to questions concerning the model that must be resolved experimentally, there remain problems of a theoretical nature. For instance, in the example we selected Ω so that the range space of $\omega(\cdot) \in \Omega$ was the finite set $\Omega = \{0, 1, 2\}$. Although the results are interpreted in terms of average behavior (because of the infinite interval and the resulting periodicity assumption), this choice of Ω corresponds, essentially, to defining deterministic sampling strategies. Moreover, this choice of Ω introduces mathematical complexities since ω does not take on a continuum of values. A characterization of Ω that eliminates this type of complexity and, at the same time, allows for a direct probabilistic interpretation of the optimal sampling strategy ω^* is possible. This, and other, descriptions of the sampling class warrant investigation.

Finally, there is the question of computing algorithms for solving the optimal sampling problem. It seems reasonably clear that the numerical search technique used in the example will prove much too tedious for more complex problems. The investigation of more efficient algorithms thus remains a prime area for research.

REFERENCES

1. Baron, S.; and Kleinman, D. L.: The Human as an Optimal Controller and Information Processor. BBN Rept. No. 1571, Dec. 1967.
2. Kalman, R. E.: Contributions to the Theory of Optimal Control. Bol. Soc. Mat. Mexicana, 1960, pp. 102-119.
3. Wonham, W. M.: Lecture Notes on Stochastic Control. Brown University Center for Dynamical Systems, Notes 67-2; 1967.
4. Kalman, R. E.; and Bucy, R. S.: New Results in Linear Filtering and Prediction Theory. J. Basic Eng., (ASME Trans.), vol. 83, 1961, pp. 95-108.
5. Elkind, J. I.; Falb, P. L.; Kleinman, D.; and Levison, W. H.: Optimal Control Theory and Design of Manual Control Systems: A Method for Predicting Control Characteristics and Display Requirements. BBN Rept. No. 1559; Oct. 1967
6. Athans, M.; and Schweppe, F. C.: On Optimal Waveform Design via Control Theoretic Concepts. Information and Control, vol. 10, no. 4, 1967, pp. 335-377.

7. Meier, L.; Peschon, J.; and Dressler, R. M.: Optimal Control of Measurement Subsystems. JACC Proc., Univ. of Pennsylvania, June 1967, pp. 527-537.
8. Athans, M.: The Matrix Minimum Principle. Information and Control, Apr. 1968.
9. Senders, J. W.; Elkind, J. I.; Grignetti, M. C.; and Smallwood, R.: An Investigation of the Visual Sampling Behavior of Human Observers. NASA Rept. No. CR-434, Apr. 1966.
10. Smallwood, R.: Internal Models and the Human Instrument Monitor. Symposium on Human Factors in Electronics, IEEE, Minneapolis, Minn., May 1966.

3. Mean Square Estimation of Human Pilot Transfer Functions

*Richard F. Whitbeck and Frederick D. Newell
Cornell Aeronautical Laboratory, Inc.*

Mean square minimization (Wiener-Hopf) methods can be used as a complement to the remnant approach in the estimation of human pilot transfer functions. Moreover, the Wiener-Hopf approach is useful in that it offers a mathematically convenient tool for extending the definition of the human pilot transfer function to multiaxis input-output tasks.

INTRODUCTION

The Cornell Aeronautical Laboratory, Inc. recently completed a project in which estimates of the transfer function of a human pilot were a primary output (ref. 1). As a preliminary to this program, a brief review of the mathematical theory, on which these estimates are based, was initiated. Basically, the pilot transfer function was found to be defined, in the open-loop case, as the linear and physically realizable impulse response that minimizes the mean square error, with this error being defined as the difference between the true pilot output and the output of his linear approximation (ref. 2, p. 25).

For the closed-loop case, an explicit definition in terms of a formal minimization does not appear to exist. Instead, one finds formulations in which the pilot is represented as the sum of a linear transfer function and a remnant. The remnant is such that the output of the linear approximation and the remnant add up to the true pilot output.

The purpose of the paper is to relate the mean square minimization concept, which seems to be the basic foundation for estimating the human pilot transfer function in the open-loop case, to the remnant concept that is presented for the closed-loop case. There are two good reasons for doing this:

(1) The mean square minimization (Wiener-Hopf) method complements the remnant approach in that it helps to establish, rather rigorously, the ground rules under which the remnant concept is valid.

(2) The Wiener-Hopf approach can be extended, in a relatively routine manner, to encompass multiaxis input-output tasks.

In the sections that follow, the mean square minimization approach will be related to the remnant concept and several of the implications will be discussed. After this, the formal procedure for extending the approach to more complex situations will be given. In general, detailed discussions of the problems that arise are not discussed. Readers who are interested in a "more depth type of coverage" are referred to reference 3.

THE RELATIONSHIP BETWEEN THE WIENER-HOPF SOLUTION AND THE REMNANT CONCEPT—OPEN-LOOP CASE.—Consider the unknown element of figure 1 which has a wide sense stationary input $x(t)$ and a wide sense stationary output $y(t)$. Internal to the system, the output of an unknown linear filter is summed with the remnant R to produce the output $y(t)$. For convenience, the figure is drawn in terms of the Laplace transform of $x(t)$, $y(t)$, etc.

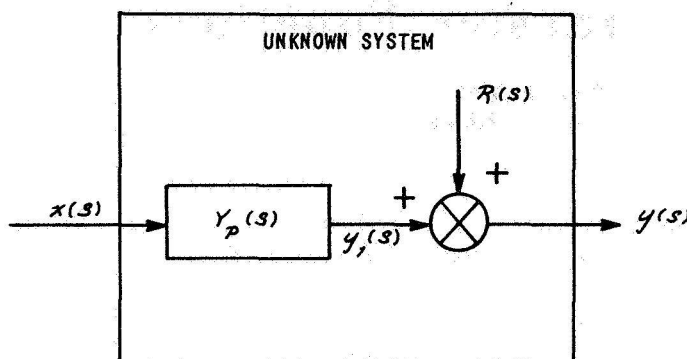


Figure 1. —Open-loop configuration.

Because $y=y_1+R$ and $xy=xy_1+xR$, it is apparent that

$$\phi_{xy} = \phi_{xy_1} + \phi_{xR} \quad (1)$$

where ϕ_{xy} = cross spectral density of x and y

ϕ_{xR} = cross spectral density of x and R

ϕ_{xy_1} = cross spectral density of x and y_1

But $Y_p(s) = \phi_{xy_1} / \phi_{xx}$, therefore

$$\phi_{xx}(s)Y_p(s) - \phi_{xy}(s) = -\phi_{xR}(s) \quad (2)$$

In equation (2), ϕ_{xx} and ϕ_{xy} are, ideally at least, spectral densities that can be measured (that is, the required data is measurable and thus the spectral densities can be computed). The difficulty is that ϕ_{xR} cannot be computed because R is not measurable. How then, can one obtain a measure of $Y_p(s)$? If we assume that the remnant is uncorrelated with the input, then the answer is straightforward, because $\phi_{xR} \equiv 0$. Therefore

$$Y_p(s) = \frac{\phi_{xy}(s)}{\phi_{xx}(s)} \quad (3)$$

if x and R are uncorrelated.

Suppose now that we are unlucky enough to have used a wide sense stationary input for which R is correlated, to some unknown extent, with x . How then, can an estimate of Y_p be obtained? One might try a formal mean square minimization and ask for a Y_p that minimizes the expected value of the mean squared remnant. In this event, the following performance index can be defined:

$$E \{J\} = E \left\{ \int_0^\infty R(t)^2 dt \right\} \quad (4)$$

Applying Parseval's theorem gives

$$E \{J\} = E \left\{ \frac{1}{2\pi j} \int_{-j}^j R(-s)R(s)ds \right\} \quad (5)$$

The integral of equation (5), call it Φ , can be rewritten as

$$\Phi(s) = \left[y(-s) - Y_p(s)x(-s) \right] \left[y(s) - Y_p(s)x(s) \right] \quad (6)$$

A well known condition (ref. 4) for finding the $Y_p(s)$ that minimizes equation (4) is then

$$E \left\{ \frac{\partial \Phi(s)}{\partial Y_p(-s)} \right\} = Z(s) \quad (7)$$

where $Z(s)$ is unknown, except for the fact that it is a function that must be analytic in the left half plane. That is, equation (7) defines a Wiener-Hopf equation. Carrying out the operation indicated in equation (7) gives:

$$\frac{\partial \Phi}{\partial Y_p(-s)} = x(-s)x(s)Y_p(s) - x(-s)x(s) \quad (8)$$

Because the $E \left\{ x(-s)x(s) \right\} \triangleq \phi_{xx}$ and $E \left\{ x(-s)y(s) \right\} \triangleq \phi_{xy}$ (ref. 5), equation (7) becomes

$$\phi_{xx}(s)Y_p(s) - \phi_{xy}(s) = Z(s) \quad (9)$$

To solve equation (9), the trick is to pick $Y_p(s)$ so that it is physically realizable, insert it into equation (9) and find that the $Z(s)$ that results gives rise to a function that has no left half plane poles.

Suppose that one now takes ϕ_{xx} , which is an even function of s (i. e. , $\phi_{xx}(s) = \phi_{xx}(-s)$), and spectrally decomposes it into the product of a left half plane component and a right half plane component. That is, write

$$\phi_{xx}(s) = \phi_{xx}^+(s) \cdot \phi_{xx}^-(s) \quad (10)$$

For example, if

$$\phi_{xx} = \frac{1}{-s^2 + 4}$$

then

$$\phi_{xx}^+ = \frac{1}{s+2}$$

and

$$\phi_{xx}^- = \frac{1}{-s+2}$$

Next perform a partial fraction expansion on ϕ_{xy}/ϕ_{xx}^- (or some other similar operation) to obtain

$$\left[\frac{\phi_{xy}}{\phi_{xx}^-} \right] = \left[\frac{\phi_{xy}}{\phi_{xx}^-} \right]_+ + \left[\frac{\phi_{xy}}{\phi_{xx}^-} \right]_- \quad (11)$$

In equation (11), $\left[\phi_{xy}/\phi_{xx}^- \right]_+$ is the sum of all the components of the partial fraction expansion that have poles in the left half plane while $\left[\phi_{xy}/\phi_{xx}^- \right]_-$ is the sum of all the components that have poles in the right half plane.

With these definitions, it is possible to verify that the correct solution to equation (9) is (see, for example, ref. 4).

$$Y_p(s) = \frac{1}{\phi_{xx}^+} \left[\frac{\phi_{xy}}{\phi_{xx}^-} \right]_+ \quad (12)$$

Note that equation (2), derived using the remnant concept, is identical with equation (9) if $Z(s)$ is set equal to $-\phi_{xR}$. If, in equation (9), $Z(s)$ is arbitrarily set equal to zero, the solution that results is

$$Y_p = \frac{\phi_{xy}}{\phi_{xx}} \quad (13)$$

and is called the particular solution of the Wiener-Hopf equation. It is also called "the solution without regard to physical realizability." In communication theory, this type of

solution occurs if infinite time delays are permitted. Illustrative examples are given in reference 3 to demonstrate that the general solution of the Wiener-Hopf equation reduces to the particular solution if the remnant is zero (the unknown system is really linear) or if the remnant is uncorrelated with the input ($\phi_{xR}=0$). The same reference also shows that if

ϕ_{xR} is required to be zero, the Y_p that results may not be physically realizable.

To conclude the open-loop case, it has been shown that the remnant concept and a formal mean square minimization can be related. It was shown that the particular solution to the Wiener-Hopf equation is identical to the generally accepted equation for estimating the human pilot's transfer function in the open-loop case.

CLOSED-LOOP ANALYSIS—Suppose now that the unknown system is enclosed in a control loop (refer to fig. 2).

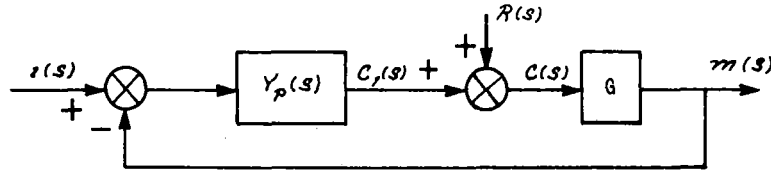


Figure 2. — Closed-loop configuration.

Assume that the input i is a wide sense stationary signal and C is still wide sense stationary after passing through the unknown element. Consider first the remnant approach.

From figure 2, write

$$C = \frac{Y_p}{1 + Y_p G} i + \frac{R}{1 + Y_p G}$$

or

$$iC = \frac{Y_p}{1 + Y_p G} i i + \frac{iR}{1 + Y_p G}$$

Therefore

$$\phi_{iC} = \frac{Y_p}{1 + Y_p G} \phi_{ii} + \frac{\phi_{iR}}{1 + Y_p G}$$

Rearranging,

$$\phi_{ii} \frac{Y_p}{1 + Y_p G} - \phi_{iC} = \frac{\phi_{iR}}{1 + Y_p G} \quad (14)$$

If it is assumed that the input (i) and the remnant (R) are uncorrelated, then $\phi_{iR}=0$ and

$$\frac{Y_p}{1+Y_p G} = \frac{\phi_{iC}}{\phi_{ii}} \quad (15)$$

Equation (15) is sometimes used to define the human pilot transfer function.

From figure 2, one may also write

$$e = \frac{i}{1+Y_p G} + \frac{G}{1+Y_p G} R$$

or

$$ie = \frac{ii}{1+Y_p G} + \frac{GiR}{1+Y_p G}$$

Therefore,

$$\phi_{ie} = \frac{\phi_{ii}}{1+Y_p G} + \frac{G}{1+Y_p G} \phi_{iR}$$

Rearranging gives

$$\phi_{ii} \cdot \frac{1}{1+Y_p G} - \phi_{ie} = - \frac{G}{1+Y_p G} \phi_{iR} \quad (16)$$

Again assuming that $\phi_{iR}=0$, gives

$$\frac{1}{1+Y_p G} = \frac{\phi_{ie}}{\phi_{ii}} \quad (17)$$

which is recognized as being another accepted equation for defining Y_p . The ratio of equation (15) to equation (16) gives another familiar equation.

$$Y_p = \frac{\frac{Y_p}{1+Y_p G}}{\frac{1}{1+Y_p G}} = \frac{\phi_{iC}}{\phi_{ie}} \quad (18)$$

Equation (18) depends on i and R being uncorrelated and it is considered to be the most appropriate equation for defining Y_p .

Turning to the Wiener-Hopf approach, questions which lead to results that are analogous to equation (14) and equation (16) may be posed. Referring to figure 2, the following problem may be formulated: minimize the mean square value of that component of C that is caused by a remnant input. That is, we are requiring something that is

appealing on physical grounds—namely, minimize the effect of remnant on the output C. In transform notation, if it is agreed that C_R is the component of C caused by the remnant, one may write:

$$C_R = \frac{R}{1+Y_p G}$$

The performance index is

$$E \left\{ \frac{1}{2\pi j} \int_{-j}^{j} \left(\frac{R(-s)}{1+Y_p(-s)G(-s)} \right) \left(\frac{R(s)}{1+Y_p(s)G(s)} \right) ds \right\} \quad (19)$$

Using figure 2, the integrand of equation (20), call it Φ , can be rewritten in terms of C and i:

$$\Phi = \left[C(-s) - \frac{Y_p(-s) i(-s)}{1+Y_p(-s)G(-s)} \right] \left[C(s) - \frac{Y_p(s) i(s)}{1+Y_p(s)G(s)} \right] \quad (20)$$

Minimize with respect to $Y_p/1+Y_p G$. Therefore

$$\partial \left(\frac{\partial \Phi}{Y_p(-s) [1+Y_p(-s)G(-s)]} \right) = -i(-s) \left[C(s) - \frac{Y_p(s) i(s)}{1+Y_p(s)G(s)} \right]$$

Taking the expected value gives the Wiener-Hopf equation

$$\phi_{ii} \frac{Y_p}{1+Y_p G} - \phi_{iC} = Z(s) \quad (21)$$

since $E \left\{ i(-s) i(s) \right\} \triangleq \phi_{ii}$ and $E \left\{ i(-s) C(s) \right\} \triangleq \phi_{iC}$.

If we set $Z(s) = -\phi_{iR}/1+Y_p G$, equation (21) is identical to equation (14), which was derived using the remnant concept. As in the open-loop case, the particular solution of the Wiener-Hopf equation (i. e., the one that results on setting $Z(s) = 0$) is

$$\frac{Y_p}{1+Y_p G} = \frac{\phi_{iC}}{\phi_{ii}} \quad (22)$$

and is equivalent to forcing i and R to be uncorrelated (compare eq. (22) with eq. (15)).

The general solution to equation (22) is

$$\frac{Y_p}{1+Y_p G} = \frac{1}{\phi_{ii}} \left[\frac{\phi_{ie}}{+} - \frac{\phi_{ii}}{-} \right]_+ \quad (23)$$

The interested reader may again refer to reference 3 where illustrative examples show that arbitrarily setting $Z(s)=0$ may lead to physically unrealizable impulse responses.

Next, what is the minimization problem which leads to the equivalent of equation (16)? Refer again to figure 2. This time, we will minimize the mean square value of that component of m that is attributable to the remnant. This component, in transform form, is

$$M_R(s) = \frac{GR}{1+Y_p G} \quad (24)$$

and the performance index is

$$E \left\{ \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left[\frac{G(-s)R(-s)}{1+G(-s)Y_p(-s)} \right] \left[\frac{G(s)R(s)}{1+G(s)Y_p(s)} \right] ds \right\} \quad (25)$$

The integrand of equation (25) can be rewritten, using figure 2, in terms of i and e :

$$\Phi = \frac{i(-s)}{1+Y_p(-s)G(-s)} - e(-s) \quad \frac{i(s)}{1+G(s)Y_p(s)} - e(s) \quad (26)$$

Minimize with respect to $\frac{1}{1+GY_p}$:

$$\frac{\partial \Phi}{\partial \left(\frac{1}{1-G(-s)Y_p(-s)} \right)} = \frac{i(s)i(-s)}{1+G(s)Y_p(s)} - i(-s)e(s) \quad (27)$$

$$\therefore E \left\{ \frac{\partial \Phi}{\partial \left(\frac{1}{1-G(-s)Y_p(-s)} \right)} \right\} = \phi_{ii} \frac{1}{1+GY_p} - \phi_{ie} = Z(s)$$

Compare equation (27) with equation (16).

The general solution to equation (27) is

$$\frac{1}{1+GY_p} = \frac{1}{\phi_{ii}} \left[\frac{\phi_{ie}}{-\phi_{ii}} \right]_+ \quad (28)$$

and the particular solution is

$$\frac{1}{1+GY_p} = \frac{\phi_{ie}}{\phi_{ii}} \quad (29)$$

As was the case in the remnant approach, the ratio of the two particular solutions of the Wiener-Hopf equation (eq. (22) and eq. (29)) gives

$$Y_p = \frac{\phi_{iC}}{\phi_{ie}} \quad (30)$$

while the ratio of the general solutions gives

$$Y_p = \frac{\begin{bmatrix} \phi_{1c} \\ \phi_{1i}^- \end{bmatrix}_+}{\begin{bmatrix} \phi_{1e} \\ \phi_{1i}^- \end{bmatrix}_+} \quad (31)$$

EXTENSION TO MULTIAXIS TASKS.—This section discusses the second point mentioned in the introduction—namely that the Wiener-Hopf approach is a useful one to take because it extends so naturally from the scalar situation to the vector (or matrix) case. In the interest of brevity, we will demonstrate the technique for a particular case, after which the methods to be followed for other situations should be relatively clear. Consider the open-loop situation depicted in figure 3.

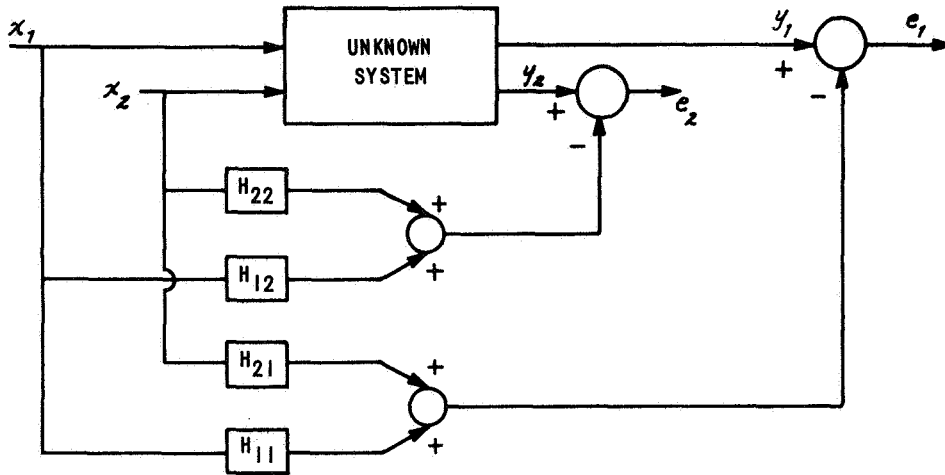


Figure 3.

In figure 3, a situation has been depicted in which one desires to find the linear transfer functions H_{11} , H_{12} , H_{21} , and H_{22} which approximate, in a mean square error sense, an unknown system that has two inputs and two outputs. That is, the performance index is

$$E \left\{ \int_0^\infty e'(t) e(t) dt \right\} = E \left\{ \frac{1}{2\pi j} \int_{-j}^j e_* e ds \right\} \quad (32)$$

In equation (32), $e(t)$ is a vector and $e'(t)$ is the transpose of $e(t)$. In a like manner,

$$e(s) = L[e(t)] \text{ and } e_* = e^*(-s)$$

Using figure 3, write the vector relationship

$$e = y - Hx \quad (33)$$

so that

$$e_* = y_* - x_* H_*$$

The integrand in equation (32) is

$$\Phi = e_*^* e = \begin{bmatrix} y_*^* & -x_*^* H_*^* \end{bmatrix} \begin{bmatrix} y - Hx \end{bmatrix} \quad (34)$$

Minimize with respect to H_* , which is a matrix:

$$\begin{aligned} \frac{\partial \Phi}{\partial H_*} &= -x_* \begin{bmatrix} y - Hx \end{bmatrix} \\ E \left\{ \frac{\partial \Phi}{\partial H_*} \right\} &= E \left\{ x_* Hx - x_* y \right\} = Z(s) \end{aligned} \quad (35)^1$$

or

$$\phi_{x_* x}^H(s) - \phi_{x_* y} = Z(s) \quad (36)$$

More explicitly,

$$\begin{bmatrix} \phi_{x_1 x_1} & \phi_{x_1 x_2} \\ \phi_{x_2 x_1} & \phi_{x_2 x_2} \end{bmatrix} \begin{bmatrix} H_{11}(s) & H_{12}(s) \\ H_{21}(s) & H_{22}(s) \end{bmatrix} - \begin{bmatrix} \phi_{x_1 y_1} & \phi_{x_1 y_2} \\ \phi_{x_2 y_1} & \phi_{x_2 y_2} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \quad (37)$$

Methods for solving matrix Wiener-Hopf equations are available in the literature. If we set the Z matrix in equation (36) equal to zero, the particular solution is

$$H(s) = \phi_{x_* x}^{-1} \phi_{x_* y} \quad (38)$$

For different situations, the procedure is generally the same—namely, to find the appropriate Wiener-Hopf equation (whether it is a scalar, vector or matrix equation), rewrite the integrand of the transformed performance index in terms of the unknowns, and take the partial with respect to the conjugate of the unknowns.

¹The operations indicated here are not defined because H_* is a matrix. However, the same result is obtained if one thinks of taking partials first with respect to the column vectors of H_* (this is a defined operation) and then "constructing" Z vectors to form the Z matrix.

REFERENCES

1. Newell, F.D.: Human Transfer Characteristics in Flight and Ground Simulation for the Roll Tracking Task. AFFDL-TR-67-30, CAL Rept. TE-2256-F-1, Jan. 1968.
2. McRuer, D.T.; and Krendel, E.S.: Dynamic Response of Human Operators. WADC-TR-56-524, ASTIA Document NR AD-110693, Oct. 1957.
3. Whitbeck, R.F.: Several Equations for the Mean Square Estimation of Human Pilot Transfer Functions. CAL FDM No. 390.
4. Chang, S.S.L.: Synthesis of Optimum Control Systems. McGraw-Hill Book Co., Inc., 1961.
5. Widrow, B.; et al.: Adaptive Antenna Systems. Proc. IEEE, No. 12, Dec. 1967.

4. Multimodality Pilot Model for Visual and Motion Cues*

*R. L. Stapleford
Systems Technology, Inc.*

A preliminary model for pilot control behavior when utilizing both visual and motion cues is presented. The model is based on a detailed review of all known data, including some unpublished data from recent experiments conducted by NASA-ARC and STI. Models for the three feedback paths (visual, angular motion, and linear motion) and integration of the various feedbacks are described.

INTRODUCTION

This paper presents an initial hypothesis for a multimodality pilot model, i.e., an approximate mathematical description of pilot control behavior when utilizing both visual and motion cues. Because there are many gaps in the currently available experimental data, certain portions of the model can be only vaguely defined. Future experiments may provide the basis for subsequent refinements and modifications. In the meantime, this preliminary model is meant to serve as a working hypothesis.

The primary motion-sensing mechanisms in a normal human are the vestibular organs—the semicircular canals and the utricles (ref. 1). The semicircular canals respond to angular and the utricles to linear accelerations of the head. There are three essentially orthogonal canals on each side of the head, so that angular accelerations with any direction can conceivably be sensed. The two utricles, on the other hand, are approximately coplanar and thus respond only to a component of the total linear acceleration.

The basic structure of the multimodality pilot model consists of three parallel feedback paths as shown in figure 1. This figure is drawn for a compensatory attitude control task, i.e., visual display is attitude error. Modifications to the visual path for other tasks or displays are obvious. It is recognized that the three noninteracting feedback paths are a gross simplification. Interactions between the visual and vestibular systems occur at several levels. At one level are the compensatory eye motions produced by the vestibular system when the head is moved. At another level are the illusions and disorientations resulting from conflicting visual and vestibular sensations. However, for the purposes of this paper the simplified model is adequate.

*This paper includes research efforts supported by the Ames Research Center, NASA, Moffett Field, California, under Contract NAS2-3650.

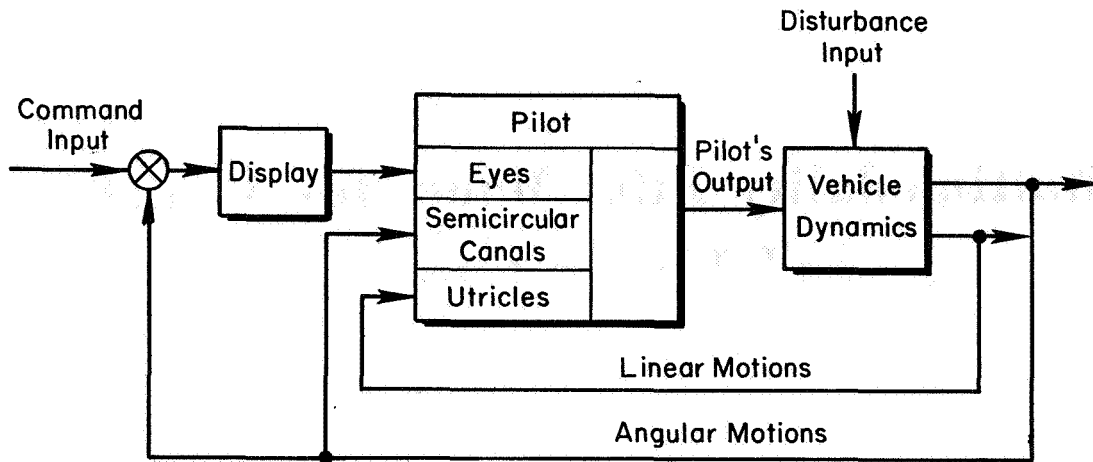


Figure 1. —Basic structure of multimodality pilot model.

The characteristics of the individual pathways of figure 1 are discussed in the next section. Later, the integration of the three feedbacks is considered. From an examination of the available data on the effects of motion cues, the conditions under which the various feedbacks are utilized are hypothesized.

CHARACTERISTICS OF INDIVIDUAL PATHWAYS

VISUAL. —Characteristics of the visual path are well known. A quasi-linear model for control tasks involving only visual cues is described in detail in reference 2. This description includes a describing function model form and adjustment rules for selecting the variable parameters. Possible modifications of the visual path due to the presence of motion cues will be discussed later.

SEMICIRCULAR CANALS. —While the semicircular canals are basically responsive to angular accelerations, their dynamic characteristics are such that over the range of frequencies normally used in manual control they can be considered as rate gyros that provide the pilot with a subjective impression of angular velocity. The model for the semicircular canal path can be represented by the elements shown in figure 2. The sensor is comprised of the semicircular canals which provide the subjective angular velocity. As in the case of visual feedback it is assumed that the pilot can provide some equalization on the sensed quantity. The final element shown in figure 2 is a lag that includes the net effects of any central processing, transmission, and neuromuscular lags. Portions of this lag are common to all three feedback paths. The characteristics of each of the three elements in figure 2 will now be discussed.

A great deal of research has been done on the sensory characteristics of the semicircular canals. A thorough summary of this work is given in reference 1. A model of semicircular canal dynamics that is particularly convenient for application to control system analyses is that shown in figure 3. Numerous experiments have demonstrated that the

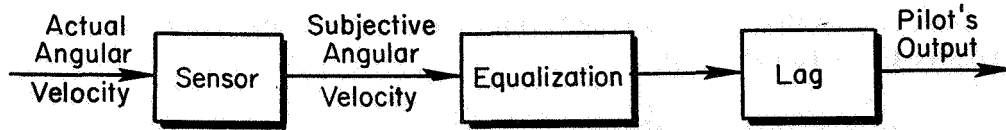


Figure 2.—Elements of the semicircular canal path.

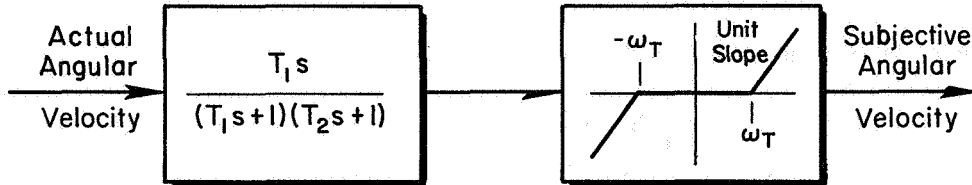


Figure 3.—Semicircular canal dynamics.

numerical values for the parameters in figure 3 have considerable intersubject variability and are a function of the axis of rotation (ref. 1). Many of these experiments have dealt with the larger time constant, T_1 . From this work the estimated values for pilots are (ref. 1):

$$T_1 \doteq \begin{cases} 6.5 \text{ sec for roll} \\ 5.3 \text{ sec for pitch} \\ 8.0 \text{ sec for yaw} \end{cases} \quad (1)$$

Data for the shorter time constant, T_2 , are much more limited. The best estimate is (ref. 1).

$$T_2 \doteq 0.1 \text{ sec for any axis} \quad (2)$$

Thus, the linear element in figure 3 acts like a bandpass filter with low-frequency cutoff at 0.125 to 0.15 rad/sec and high-frequency cutoff at 10 rad/sec, and with nearly unity gain over the frequency range 0.3 to 5 rad/sec. Over this range the semicircular canals function as a rate sensor.

Numerical values for the threshold shown in figure 3 are based on experiments to determine the minimum detectable constant angular acceleration or step velocity change. For a step acceleration input of magnitude α , the output of the linear element in figure 3 would be

$$T_1 \alpha \left[1 - \left(\frac{T_1}{T_1 - T_2} \right) e^{-t/T_1} + \left(\frac{T_2}{T_1 - T_2} \right) e^{-t/T_2} \right]$$

The steady-state response is then $T_1\alpha$. The minimum detectable angular acceleration, α_{\min} , is then related to the threshold by

$$T_1\alpha_{\min} = \omega_T \quad (3)$$

Using the α_{\min} values from reference 3, i.e.,

$$\alpha_{\min} = \begin{cases} 0.5 \text{ deg/sec}^2 & \text{for roll or pitch} \\ 0.14 \text{ deg/sec}^2 & \text{for yaw} \end{cases} \quad (4)$$

and the time constants from equation 1, the threshold is estimated to be

$$\omega_T = \begin{cases} 3.2 \text{ deg/sec} & \text{for roll} \\ 2.6 \text{ deg/sec} & \text{for pitch} \\ 1.1 \text{ deg/sec} & \text{for yaw} \end{cases} \quad (5)$$

As a matter of interest, these thresholds are an order of magnitude greater than those that would be specified for flight control system rate gyros.

The above expression can also be used to approximate the latency time (time to detect an input acceleration), T_L , by

$$\begin{aligned} \omega_T &= T_1\alpha \left[1 - \left(\frac{T_1}{T_1 - T_2} \right) e^{-T_L/T_1} + \left(\frac{T_2}{T_1 - T_2} \right) e^{-T_L/T_2} \right] \\ &\doteq T_1\alpha \left(1 - e^{-T_L/T_1} \right) \end{aligned} \quad (6)$$

Equation 6 has been shown to match measured latency times quite accurately, e.g., reference 3.

Values for the threshold can also be estimated from the minimum detectable step change in velocity. For a step velocity change of magnitude, ω , the output of the linear element of figure 3 is

$$\frac{T_1\omega}{T_1 - T_2} \left(e^{-t/T_1} - e^{-t/T_2} \right)$$

The maximum value of the output is easily shown (using $T_1 \gg T_2$) to be approximately ω . Therefore the threshold, ω_T , is also approximately equal to the minimum detectable step velocity change. Threshold estimates obtained in this manner are in rough agreement with those given in equation 5, although there is considerable intersubject variability. (See ref. 1 for a more complete discussion of this subject.)

In most cases of manual vehicular control, the motions are considerably above the thresholds noted above and the primary concern is in the frequency range of 1 to 5 rad/sec. Then the sensor dynamics for the semicircular canal path are adequately approximated by

$$\frac{\text{subjective angular velocity}}{\text{actual angular velocity}} = \frac{1}{T_2 s + 1} = e^{-T_2 s} \quad (7)$$

* The other two elements in the semicircular canal path are equalization and lag. Unfortunately, there are no direct data on these two and there is only a limited amount of inferential data. Some of the data of reference 4 indicate that relatively large lead equalization (roughly 1 second) is possible. Whether or not the pilot can generate lead equalization as large as that measured for visual tracking is unknown. In fact, the mechanism for generating lead in visual tasks is still not completely understood. Until more concrete data can be obtained, it will be hypothesized that the lead in the semicircular path can be as large as that used in the visual path.

It is also theoretically possible for the pilot to use lag equalization in the semicircular path. However, as will be covered later, the primary function of the path appears to be to supply lead equalization. In none of the cases considered to date has lag equalization been desirable. Therefore, it is assumed that the semicircular canal equalization is of the form $K_1(T_3 s + 1)$.

As indicated above, there are no direct data on the lag element. Preliminary results from some recent experiments (ref. 5) indicate a lag of roughly 0.1 second. Completion of the detailed analyses may allow a refined estimate of the time delay, but for now a value of 0.1 second will be assumed.

Combining the above gives the model for the semicircular canal path shown in figure 4.

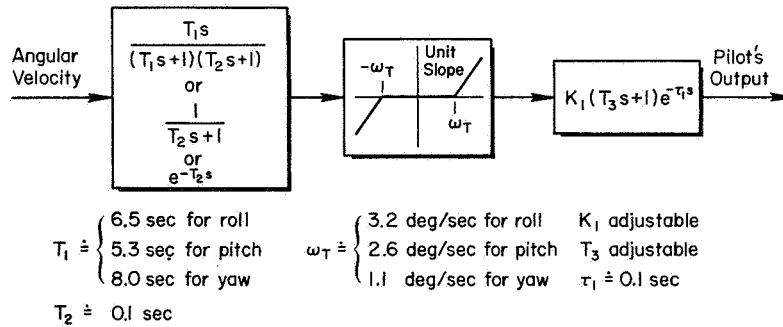


Figure 4.—Model for semicircular canal path.

UTRICLES

The model for the utricular path consists of three elements similar to those used for the semicircular path—sensor, equalization, and lag. While little data is available on the sensor dynamics of the utricles, it is widely accepted that they are sensitive to linear accelerations in the plane of the utricular maculae. This plane is inclined front end upward approximately 30° from the horizontal in the upright head.

It should be noted that, like conventional accelerometers, the utricles do not respond to inertial accelerations but to the total applied force. For example, on the ground the utricles respond to tilting the head relative to the gravity vector just as an accelerometer attached to the head would. For simplicity, the inputs to the utricles will hereafter be referred to as accelerations. However, the reader should remember that this means the accelerations that would be sensed by an accelerometer.

The most recent data on utricular sensory dynamics is that given in reference 6. That report suggests a model for the sensory dynamics of the form

$$\frac{\text{subjective acceleration}}{\text{actual acceleration}} = \frac{(T_5/T_4)(T_4s+1)}{(T_5s+1)(T_6s+1)} \quad (8)$$

where

$$T_4 \doteq 13 \text{ sec}$$

$$T_5 \doteq 5.2 \text{ sec}$$

$$T_6 \doteq 0.67 \text{ sec}$$

Over the frequency range of interest in most vehicular-control situations equation 8 can be adequately approximated by

$$\frac{\text{subject acceleration}}{\text{actual acceleration}} \doteq \frac{1}{T_6s+1} \quad (9)$$

The utricular threshold is so small, on the order of 0.01g or less, that it will have a negligible effect in most vehicular control situations. There are also very few data on latency times but reference 3 does show latency times on the order of 1 second for accelerations of 0.1g.

Information on the equalization and lag elements is limited to the reference 7 analysis of some of the data from reference 3. The results for the one case involving a utricular feedback indicated the presence of a first-order lead at 3 rad/sec and a time delay of 0.3 second. The lead is identified as pilot equalization and the time delay as the low-frequency approximation to the dynamics of the lag element. This analysis provides some evidence to support the inclusion of pilot adjustable lead in the utricular path. However, there are no data on the adjustable range of the lead or the possibility of lag equalization. The possible desirability of lag equalization cannot be ruled out because of the extremely wide range of acceleration numerator zeros that can occur in vehicle transfer functions. These zeros are strong functions of the vehicle stability and control properties and the pilot's location. It is, at least theoretically, possible that in some cases lag equalization would be helpful.

The assumed model for the utricular path is shown in figure 5.

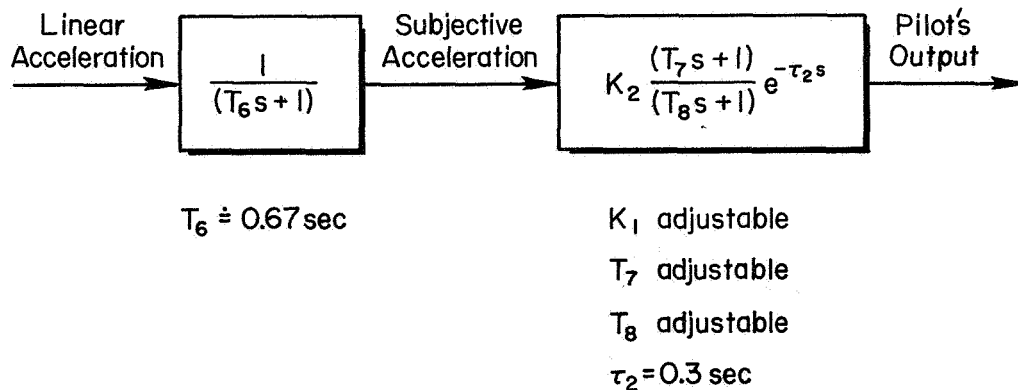


Figure 5.—Model for utricular path.

FEEDBACK INTEGRATION

The main objective here is to determine how the three feedback paths are combined and integrated by the pilot. We will start with a review of the available data and then postulate a general model. Experimental data that include pilot describing function measurements will be discussed first.

Table 1 summarizes those experiments in which pilot describing functions were measured for both fixed-base (visual-only tracking) and moving base or flight. In all cases the describing functions were measured from the displayed error to the pilot's control output. Consequently, the moving-base or flight results contain the effects of the parallel visual and motion feedbacks.

For one reason or another, the results of the first four references listed in table 1 are inconclusive when attention is centered only on motion effects. A limited amount of data is provided by references 3, 12, and 13, but Shirley (ref. 4) and NASA/STI (ref. 5) provide extensive experimental results of a highly complementary nature.

Shirley measured describing functions for 40 different controlled elements with and without motion cues. The motion cues resulted from rolling about a fixed horizontal axis with the subject's head nearly on the axis.

The NASA/STI tests used only 3 controlled elements but examined changes in the motion cues. Both roll and lateral position motions were used to simulate different head positions relative to the vehicle c.g., and a variety of washouts were tested. In addition to overall describing function measurements, similar to those made by Shirley, some unique experiments were conducted using two inputs in an attempt to measure the visual and motion feedbacks separately.

As the analyses of the NASA/STI results are not yet complete, a detailed comparison with Shirley's results cannot be made. A preliminary comparison indicates many similarities and several differences. As there were a number of dissimilarities in the experimental setups that could account for the differences in the results, we will concentrate here on the results which are in agreement. Both tests, as well as those of Meiry (ref. 3) and Newell (ref. 12 or 13), show that the major effects of motion cues are to increase pilot gain and reduce his high-frequency phase lags. The lead information supplied by the motion reduces the phase lag in the overall (combined visual and motion paths) describing function by an amount that is generally equivalent to a time delay reduction of 0.1 to 0.2 sec. As a result the pilot can, and does, increase his gain and crossover frequency to maintain similar, and frequently smaller, phase margins. The increase in crossover frequency is typically on the order of 1 rad/sec.

The two input data from the NASA/STI experiments are also providing information on changes in the visual feedback when motion cues are present. The low-frequency gain in the visual path is increased, less lead is used, and the time lag may be increased. With the lead provided by the motion cues, the pilot need not supply as much lead in the visual path and can allow his delay to increase. Because the motion feedbacks provide more lead than the visual feedback, the motion paths are dominant (have higher gain) at the higher frequencies, and the visual path dominates at lower frequencies. The two are of comparable magnitude in the region of 2 to 4 rad/sec. It also appears that the semicircular canal path is always utilized to a significant degree but that the utricular path is used only when the linear acceleration is a desirable feedback.

Newell's data (ref. 12 or 13) demonstrate the important difference between motion effects with a command or a disturbance input (see fig. 1 for the difference between a command and a disturbance input). Newell found that with a command input and a compensatory display, the in-flight describing functions were approximately the same as those for the

TABLE 1. -- DESCRIBING FUNCTION DATA

REF.	MOVING-BASE DEVICE	DISPLAY		MOVING-BASE INPUT	RESULTS	REMARKS
		FIXED- BASE	MOVING BASE			
8	Princeton Navion	CRT	CRT	Command	Flight τ 's higher by 0.1 - 0.2 sec than ground. Pilot's longitudinal gain in flight approximately 1/2 ground value.	Differences observed may not be totally due to motion differences because: 1. Different subjects for in-flight and ground data. 2. Considerable dis- tractions during in-flight tests.
9	NASA TV-2	Gyro horizon	Gyro or natural horizon	Disturbance	Inconclusive. Amplitude and phase differences were inconsistent.	Data reduction incorrect. Used error rather than input, cross-spectra.
10	Air Force Variable- Stability T-33	Gyro horizon	Natural horizon	Disturbance	Only conclusive dif- ference was lower gain on ground for frequen- cies less than approxi- mately 0.3 rad/sec	Considerable variability in the flight data.
11	MIT Any-Two- Degree of Angular Freedom Motion Simulator	CRT	CRT	Disturbance	Moving-base gains lower and lags larger than fixed-base.	Author attributes dif- ferences to a combination of "simulator nonlinear- ities, a poorly-marked and less sensitive moving- base display grid, and perhaps vestibular con- fusion and insensitivity to very small deflec- tions."

TABLE 1. -- CONCLUDED

REF.	MOVING-BASE DEVICE	DISPLAY		MOVING-BASE INPUT	RESULTS	REMARKS
		FIXED- BASE	MOVING- BASE			
3	MIT Any-Two- Degrees of Angular Freedom Motion Simulator	CRT	CRT	Disturbance	1. With visual and semi-circular canal stimulus, τ reduced 0.1 sec from visual alone. 2. With semicircular canal and utricular stimuli, τ reduced 0.2 sec and crossover frequency twice that for visual alone.	Case 1 was rotation about vertical axis. Case 2 was rotation about horizontal axis. Results further analyzed in Ref. 7.
12, 13	Air Force Variable- Stability T-33	Gyro horizon	Gyro or natural horizon	Command or disturbance	1. Small differences between ground and flight with command input. 2. Flight with disturbance input had significantly higher crossover frequencies and lower τ 's than ground.	For flight tests, gyro horizon always used with command input and natural horizon always used with disturbance input. Compensatory display used for command inputs.
4	MIT Any-Two- Degrees of Angular Freedom Motion Simulator	CRT	CRT	Disturbance	Describing functions and performance measures given for 40 controlled elements. Generally, the addition of motion significantly increased the pilot gain, reduced high frequency phase lags, and improved performance.	Motion was roll about a horizontal axis with subject's head nearly on the axis.
5	NASA-ARC Six-Degree- of-Freedom Simulator	Gyro horizon	Gyro horizon	Disturbance (some tests with 2 inputs)	1. Similar to Ref. 4 for comparable controlled elements 2. Direct measurements of visual and motion feedbacks separately.	1. Motion was roll and lateral position. 2. Simulated head position and washouts were varied.

ground tests. In hindsight, this is not surprising. With a command input and compensatory display, the sensed motions are not the same as the visual display. The pilot senses the actual motion of the aircraft but sees the difference between that motion and the input. Thus, one might expect the pilot to ignore the motion cues and operate strictly on the visual cues. On the other hand, with a disturbance input the sensed motions and the visual display agree. The increase in crossover frequency and reduction in effective time delay found by Newell in this case agree quite well with the results of references 3, 4, and 5.

Unfortunately, Newell's results cannot be conclusively attributed to motion cue effects alone as there was also a difference in the display. The ground and flight tests with a command input utilized a gyro horizon display, while for flight with the disturbance input the natural horizon was used. Newell in references 12 and 13 attributes the changes in pilot characteristics to the improved display. However, since the results do check with those obtained in three other experiments, it appears likely that most of the differences were due to the pilot's utilization of the motion cues.

A literature survey on the effects of motion clues on pilot performance was also made for any additional clues on how or when a pilot utilizes motion feedbacks. Unfortunately, we could find only five experiments that met all of the following criteria: the pilot's task involved tracking, vehicle dynamics and the display were nearly identical for the fixed-base and moving-base tests, the motion simulator did not introduce significant spurious cues, and performance data were given. The five are references 3, 4, 5, 14, and 15. The performance data, however, do not prove very enlightening. There are several apparent discrepancies among these five tests. These discrepancies may be due to differences in: the training and experience of the subjects, the visual display, the magnitude and bandwidth of the input, or the manipulator. Nevertheless, two noteworthy results were found.

The reference 14 experiments showed that a pilot can utilize motion cues with a command input if he has a pursuit display. This finding is completely compatible with the previous conclusions on display-input effects. With a pursuit display the pilot can visually perceive the same vehicle motions he senses via his vestibular system. Since the two agree he can use both.

The reference 15 experiments demonstrated the importance of the training and experience of the subjects. These tests involved the simulated roll angle and lateral position control of a hovering helicopter. Two experienced helicopter pilots could not control the vehicle fixed base, but could moving base. Yet, two of three nonpilot subjects did better fixed base than moving base.

Based on the experimental data discussed above, the following hypotheses regarding the utilization of motion cues are made:

- (1) Motion cues cannot be used and will be ignored when tracking a random-appearing¹ command input with a compensatory display.
- (2) With any input-display combination except that indicated in (1), motion cues may be utilized.
- (3) The addition of motion cues will greatly enhance performance when considerable lead equalization and/or effective-time-delay reduction in the visual feedback would be beneficial.
- (4) With motion cues the gain in the visual path may be considerably increased over what it would be for fixed-base operation, the lead may be reduced, and the effective time delay increased.

¹With a simple, predictable input the pilot may be mentally able to separate the input and the vehicle response and to operate as if he had a pursuit display.

(5) The motion feedbacks will dominate at high frequencies, the visual feedback will dominate at low frequencies, and they will be of comparable magnitude in the region of 2 to 4 rad/sec.

(6) The utricular path will not be used unless the linear acceleration feedback is more favorable than the semicircular canal feedback.

These hypotheses plus the models for the semicircular canal and utricular paths (figs. 4 and 5) form the preliminary multimodality pilot model.

REFERENCES

1. Peters, Richard A.: Dynamics of the Vestibular System and Their Relation to Motion Perception, Spatial Disorientation, and Illusions. Systems Technology, Inc., Tech. Rept. 168-1, Feb. 1968.
2. McRuer, Duane T.; Graham, Dunstan; and Krendel, Ezra S.: Manual Control of Single-Loop Systems: Parts I and II. J. Franklin Institute, vol. 283, no. 1, Jan. 1967, and no. 2, Feb. 1967.
3. Meiry, Jacob L.: The Vestibular System and Human Dynamic Space Orientation. MIT, Man-Vehicle Control Lab., Thesis No. T-65-1, June 1965.
4. Shirley, Richard S.: Motion Cues in Man-Vehicle Control. MIT, Sc.D. Thesis, Dec. 1967.
5. Young, L. R.; and Meiry, J. L.: A Revised Dynamic Otolith Model. Third Symposium on the Role of the Vestibular Organs in Space Exploration, SP-152, 1967.
6. Alex, F. R.: Analysis of Two Manually Controlled Compensatory Systems with Inclusion of Motion Cues. Systems Technology, Inc., Working Paper 168-2, Feb. 1967.
7. Seckel, Edward; Hall, Ian A. M.; McRuer, Duane T.; et al.: Human Pilot Dynamic Response in Flight and Simulator. WADC TR 57-520, Aug. 1958.
8. Kuehnel, Helmut A.: Human Pilots' Dynamic-Response Characteristics Measured in Flight and on a Nonmoving Simulator. NASA TN D-1229, Mar. 1962.
9. Smith, Harriet J.: Human Describing Functions Measured in Flight and on Simulators. Second Annual NASA-University Conference on Manual Control. NASA SP-128, 1966, pp. 279-290.
10. Kilpatrick, Philip S.: Bending Mode Acceleration Influence on Pilot Control of Flexible Booster Dynamics. MIT, Man-Vehicle Control Lab., Thesis No. T-65-2, Sept. 1965.
11. Newell, Fred D.; and Pietrzak, Paul E.: In-Flight Measurement of Human Response Characteristics. AIAA Paper No. 67-240, Feb. 1967.
12. Newell, F. D.: In-Flight and Ground Simulation Measurements of Pilot Transfer Characteristics in the Compensatory Roll Tracking Task. Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967, pp. 99-119.
13. Sadoff, Melvin; McFadden, Norman M.; and Heinle, Donovan R.: A Study of Longitudinal Control Problems at Low and Negative Damping and Stability with Emphasis on Effects of Motion Cues. NASA TN D-348, Jan. 1961.
14. Benjamin, Peter: Visual and Motion Cues in Helicopter Flight. MIT, Man-Vehicle Control Lab., Thesis No. T-66-1, Jan. 1966.

5. Synthesis of Human Response in Closed-Loop Tracking Tasks

James J. Adams
Langley Research Center, NASA

Previous investigations have determined linear, average constant-coefficient models that provided a reasonable reproduction of a human controller's response in tracking tasks. However, when such models are put in a control loop in place of a human, they do not exactly reproduce the system response obtained with the human controller. These previous studies suggest that time variations of the model coefficients and a random noise signal should be added to the linear model to obtain a more accurate representation of the human's response.

In this present study, experiments have been conducted to determine the variability in a human subject's control stick response to displayed displacement and rate of change of displacement to aid in the implementation of the time variations to be included in the model. Also, additional tracking tasks were made to obtain a definition of the characteristics of the random noise to be added to the model. These two factors were then added to the model, and this composite model was placed in a control loop in place of the pilot. The results demonstrate that this composite model reproduces time-history characteristics and mean-square system errors that more closely match the human subject than does the linear model.

INTRODUCTION

While linear models of pilots have proven to be very useful in the analysis of manually controlled systems (for example, as shown in references 1 and 2), these linear models lack the random nature of human subjects and, therefore, have lacked complete accuracy. Taking the average gains as measured by parameter tracking methods, constructing a model with these gains, and placing this model in a control loop in place of the pilot results in a lower mean-square system error than obtained with the human subject. This fact is demonstrated in reference 3 where the task involved the compensatory control of a random signal. The experiments described in this paper test the hypothesis that adding a random noise signal and the time varying gains to the linear model will provide a greatly improved match of the system error to that error obtained with a human subject.

The fact that the human controller is not strictly linear is illustrated in figure 1. The figure presents time histories obtained during tests of compensatory tracking with acceleration dynamics in which the forcing function was a pure sine wave. If the subject did respond in a linear manner the system error would be a sine wave, which, after the initial transient had died out, would be of the same frequency as the disturbance function. However, the figure clearly indicates that this is not the case. These results, and similar results presented in reference 4 where the controlled dynamics were a simple amplifier, suggest that the human controller is a noise source, and therefore his output contains a random signal. In addition to the evidence of a noisy control signal given in the figure, there is also evidence of time varying linear system characteristics. Short sections of very uniform exponential divergences, such as would be produced by an unstable or poorly damped linear system, can be seen. These divergences indicate that the human should also be considered as being a time varying controller. These two factors, the random output and the time variation, were considered as being the factors that should be added to the linear model to provide an improvement in the representation of the human subject.

TIME VARIATIONS

To determine just how much time variation and how much random signal to add to the linear model, some special tests were made. To determine the magnitude of the time variation that should be used, the static response in stick position to displayed signals was determined. In these tests the subject observed a large oscilloscope display on which, in the first series, a horizontal line moved from top to bottom or bottom to top at fixed rates. These rates were from 0 to ± 80 cm/sec, or, in terms of line-of-sight rotation, from

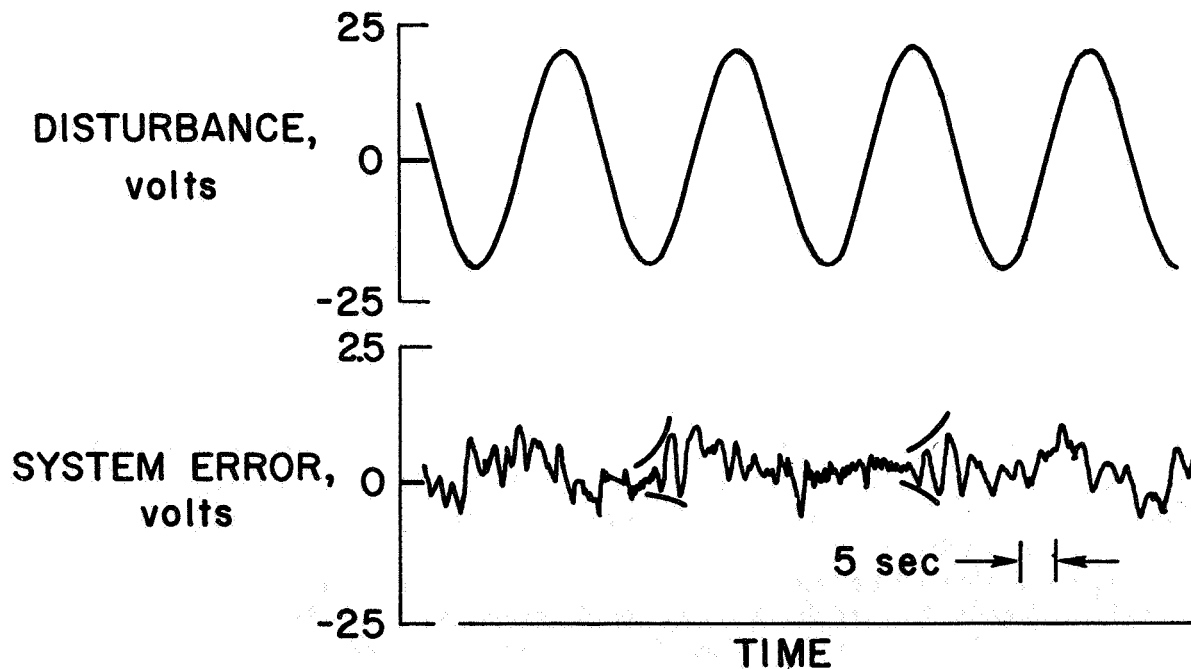


Figure 1.—Closed-loop response to a sine wave disturbance.

0 to ± 1.1 rad/sec. The subject responded by moving a side-arm controller to a position that he felt was an appropriate response and held this response long enough to establish a fixed value. These were open-loop tests. The controller deflection did not influence the display. The subjects were not instructed to establish any particular gain for these responses, but only to be consistent.

Typical results of these tests are shown in figure 2. The first thing that can be deduced from this plot is that the response is linear. A straight-line average fit to this data can be drawn which defines the subject's gain. Also, straight-line boundaries for the data scatter can be drawn, establishing the subject's spread in his gain. The percentage of this spread from the average gain for each of the subjects is shown in the figure. It can be seen that this spread averages ± 45 percent for all of the subjects. This ± 45 percent was used as the amplitude of time variation of the lead gain for the linear model. The average gain was taken as that value established by the parameter tracking method, and this average value was perturbed ± 45 percent.

Similar tests were performed to establish the spread in response to display displacement. The spread in the use of displacement was found to be ± 30 percent, and this ± 30 percent was applied to the static gain of the linear model.

The time characteristic of the gain variations was implemented as follows. Records such as that presented in figure 1 show that manually controlled systems are poorly controlled for periods of 5 seconds. To duplicate this condition the variable gains were given a saw tooth variation that had a cyclic period of 20 seconds. The static and lead gains were varied together - both high and both low at the same time. Thus the model controlled system operated at nominal gains for 10 seconds during every cycle, and low and high gains for 5 seconds.

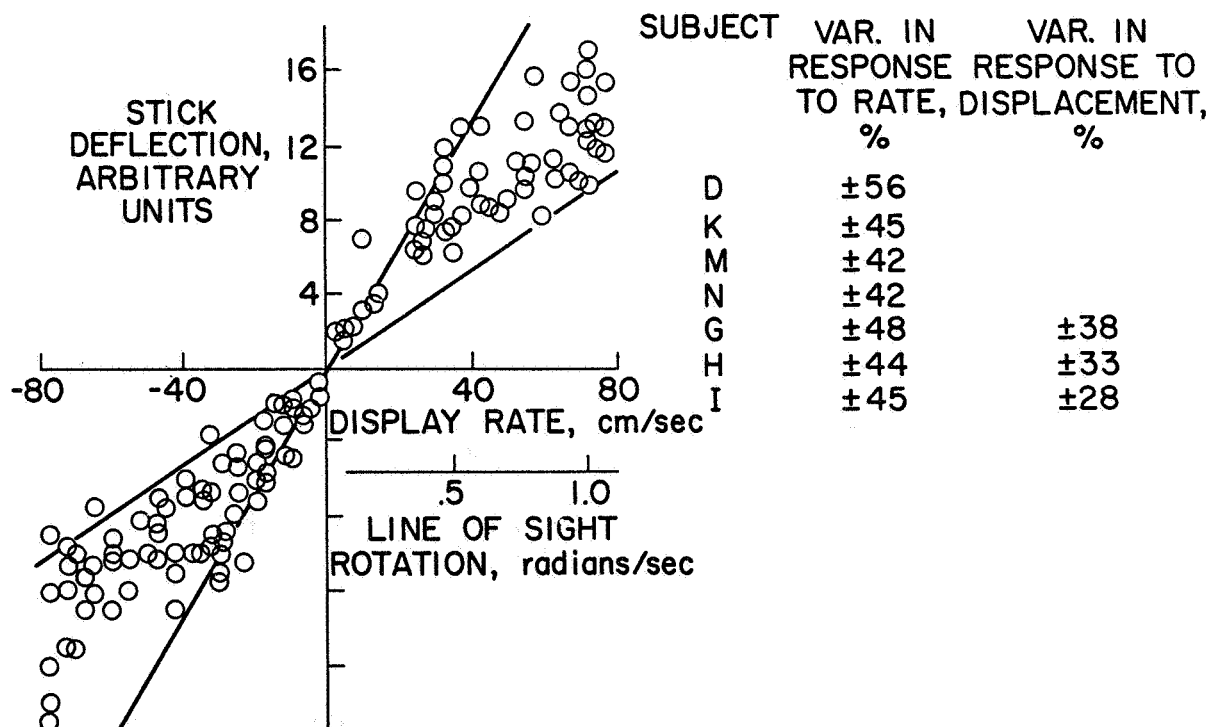


Figure 2.—Static stick response to displayed signals.

RANDOM SIGNAL

The amplitude and spectral density of the pilot-injected noise were obtained from data gathered during compensatory tracking tests. In these tests the subjects controlled acceleration dynamics and the disturbance function was adjusted to have different mean-square value from run to run. The pilot-injected noise was taken to be the difference between the pilot's output and the output of the adjustable model used in the parameter tracking method as described in reference 5.

The variance of the various signals around the control loop for one subject is shown in table 1.

TABLE 1. —VARIANCES AND RATIOS

Variance				Ratios	
Disturbance, volts ²	System error, volts ²	Stick, volts ²	Pilot noise, volts ²	$\frac{\text{Error}}{\text{Disturbance}}$	$\frac{\text{Pilot noise}}{\text{Stick}}$
5.61	3.57	0.775	0.356	0.64	0.46
15.4	16.2	1.81	.535	1.05	.30
53.1	35.7	11.0	3.36	.67	.31

Although there were variations between subjects, the values shown are typical. It can be seen that all of these values increase as the variance of the disturbance function increases. To determine a generalized nondimensional number, the ratio of the variance of the measured pilot noise to stick deflection was formed. This ratio was the most nearly constant of all the ratios that could be formed. The average value for the subject whose data are shown in table 1 is 0.36. For all the pilot subjects the values of ratio of pilot noise to stick deflection ranged from 0.20 up to 0.47.

The spectral density of the pilot noise is shown in figure 3. Also presented is the square of the input-output ratio of two first-order lags with break frequencies of 10 rad/sec. The good agreement between these two curves is clearly seen. For all the subjects the shape of the spectral density of the pilot noise had the characteristic of a second-order system with a break frequency that varied from 10 to 15 rad/sec. To simulate this pilot noise, the output of a noise generator was passed through two first-order filters with break frequencies of 10 rad/sec. The amplitude of the noise was adjusted until the ratio of stick deflection to noise was the proper value. This adjustment was done while the model was in the control loop. Since the model was in the control loop, adding noise to the output of the model changed the stick deflection in a manner that could not be accurately predicted. Therefore, the adjustment to the amplitude of the noise had to be adjusted in an iterative manner. It was possible after one iteration to arrive at approximately the desired amplitude for the added noise.

SYNTHESIS OF HUMAN RESPONSE

The simulated pilot noise and the time varying gains were added to the pilot model in a manner illustrated by the block diagram shown in figure 4. The ± 30 -percent variation on the static gain was placed at the output of the model. The ± 45 -percent total variation on the lead

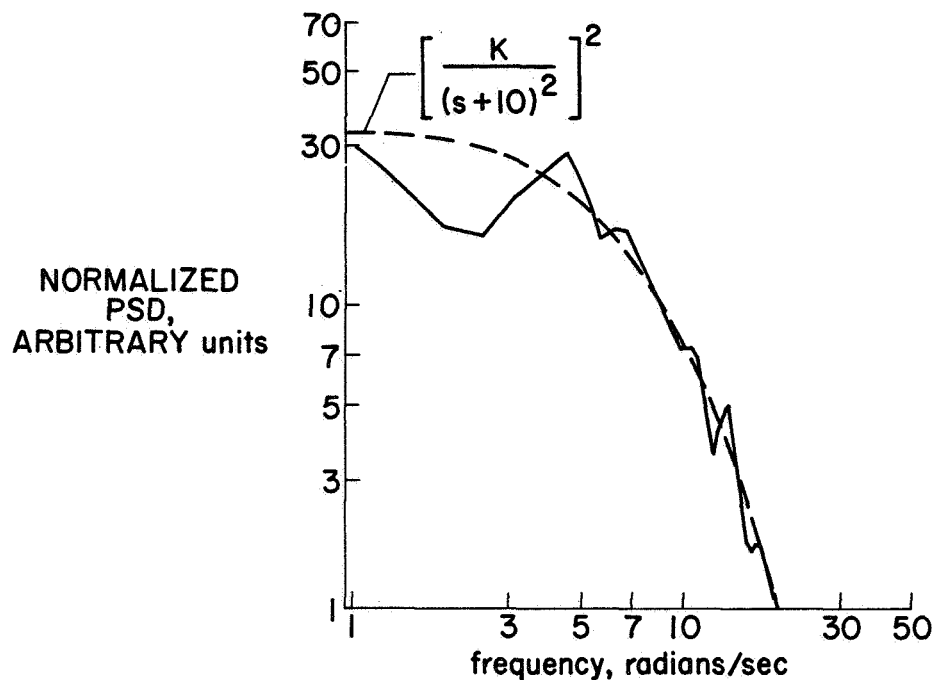


Figure 3. —Power spectral density of pilot noise.

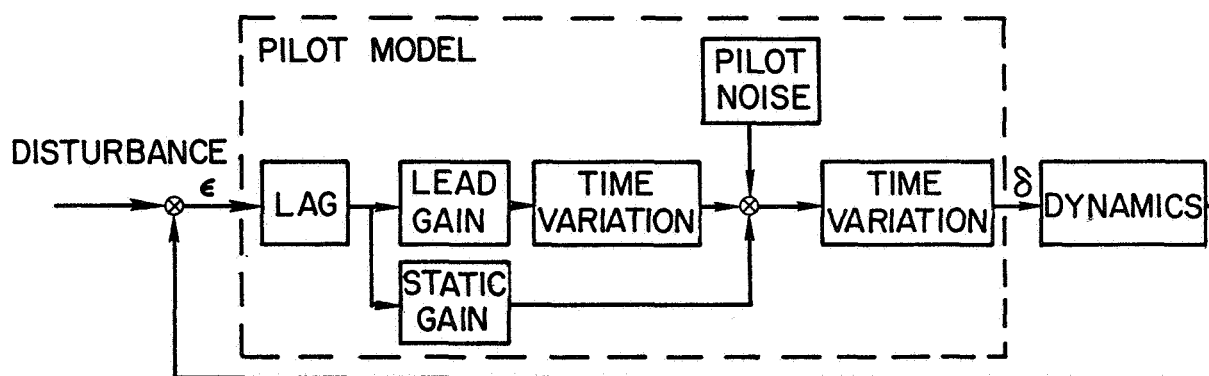


Figure 4. —Block diagram of closed-loop system with composite model.

gain was achieved by putting an additional ± 15 -percent variation on the lead gain. The pilot noise was inserted in the model ahead of the output of the model. With this arrangement the ratio between noise and stick deflection would remain constant even though the static gain was being varied.

The results obtained with this composite model are presented in figure 5. Presented are the system errors obtained with the human controller in the loop, with the composite model in the loop, and with the linear model in the loop. The same disturbance function was used in obtaining each of these three records. It can be seen that the composite model provides a great improvement in the match of the time-history characteristics of the system error over that obtained with the linear model. The ratios of the mean-square values of the

system error to disturbance are also presented to further illustrate the improvement in the match. For this particular example the ratios are as listed in table 2.

The human pilot did not perform with the same ratio of mean-square system error to disturbance for each of the 3-minute tests that he made, as illustrated in table 1. Neither did the composite model. To illustrate that the spread in performance with the composite model was the same as that obtained with the human pilot, table 3 is presented. The figure shows the average of the ratio of mean-square system error to disturbance and the variation for the

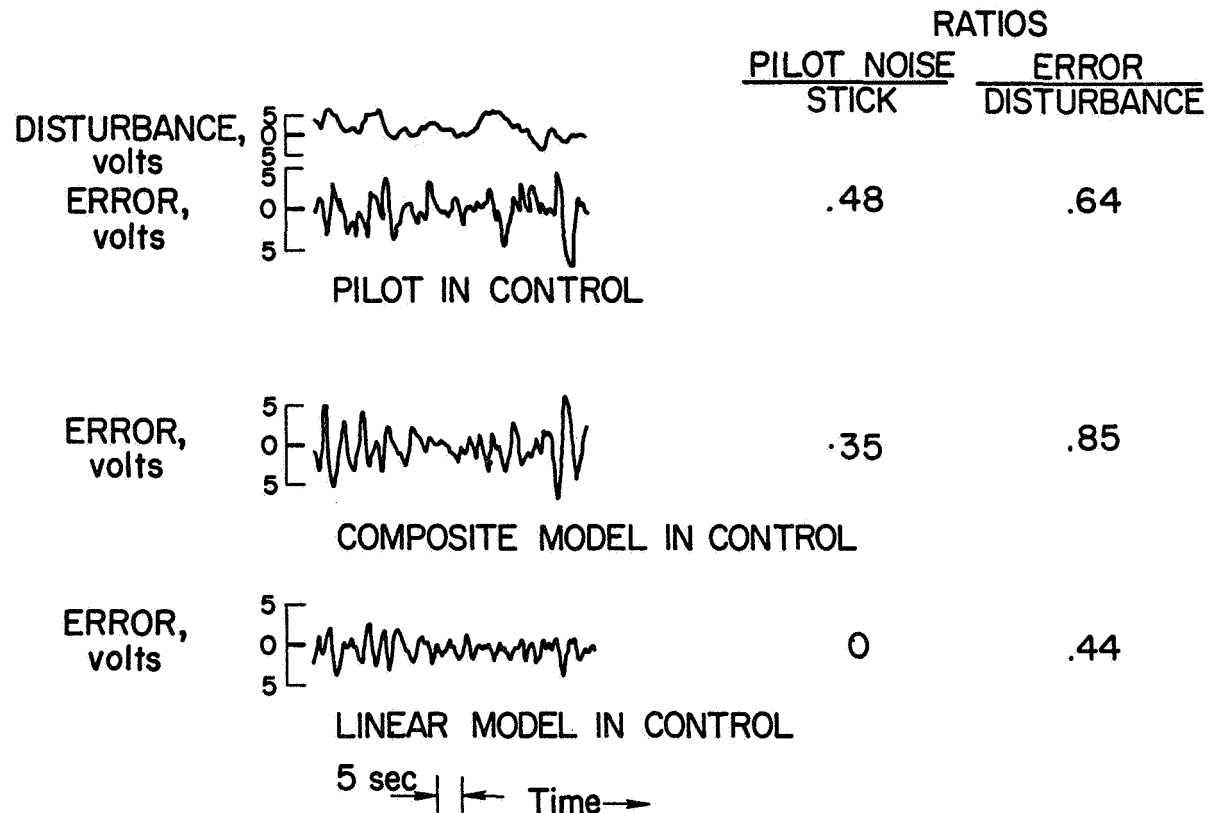


Figure 5.—Comparison of system response with pilot, composite model, and linear model in control.

TABLE 2.—DATA FOR FIGURE 5

	Noise Stick', $\frac{n^2}{\delta^2}$	Error Disturbance', $\frac{e^2}{D^2}$
Pilot	0.48	0.64
Composite model	.35	.85
Linear model	0	.44

three tests performed by the pilot, for six tests performed with the composite model, and for five tests performed with the linear model. Here again it can be seen that an improvement in the match to the human pilot is obtained with the composite model over that obtained with the linear model.

The largest part of the improvement achieved with the linear model was provided by the addition of the random noise. Table 4 illustrates this fact. It can be seen that nearly four-fifths of the improvement is provided by the addition of the noise.

TABLE 3. —AVERAGE DATA

	$\frac{\text{Error}}{\text{Disturbance}}, \frac{e^2}{D^2}$
Pilot	0.78; +0.26, -0.12
Composite model	.73; +.37, -.19
Linear model	.35; +.08, -.05

TABLE 4. —AVERAGE DATA

	$\frac{\text{Error}}{\text{Disturbance}}, \frac{e^2}{D^2}$
Pilot	0.78
Composite model	.73
Linear model with added noise only	.70
Linear model with variable gains only	.43
Linear model	.35

CONCLUSIONS

The data presented in this paper demonstrate that human pilot response can be more closely matched by a linear model to which have been added time variations in the gains and a random signal than with the linear model alone. In the case of multi-axes tasks and in cases where side tasks must be performed in addition to the closed-loop control, references 6 and 7 have prescribed additional changes that should be included in the model. These additions involve the reduction of certain gains to zero for prescribed lengths of time. Other changes in the gains of the linear portion of the model that correspond to various vehicle dynamics are presented in references 3, 8, and 9. By following the suggestions for the synthesis of pilot response presented in this paper and in the references, pilot models for a large portion of all probable control situations can be constructed. The simple, linear, constant-coefficient models can be used to predict, or confirm, gross system performance. For detailed studies of precision of control, structural loads, and fuel requirements the composite model should be used.

REFERENCES

1. Ashkenas, Irving L.; Jex, Henry R.; and McRuer, Duane T.: Pilot-Induced Oscillations: Their Cause and Analysis. Norair Rept. NOR-64-143 (Rept. STI TR-239-2), Northrop Corp., June 20, 1964.
2. Adams, James J.: Application of Human Transfer Functions to a Design Problem. Third Annual NASA-University Conference on Manual Control, NASA SP-144.
3. Adams, James J.; and Bergeron, Hugh P.: Measurements of Human Transfer Functions With Various Model Forms. NASA TN D-2394, 1964.
4. Pew, R. W.: Sine Wave Tracking Revisited. Second Annual NASA-University Conference on Manual Control, NASA SP-128.
5. Adams, James J.; and Bergeron, Hugh P.: Measured Variations in the Transfer Function of a Human Pilot in Single-Axis Tasks. NASA TN D-1952, 1963.
6. Adams, James J.; Bergeron, Hugh P.; and Hurt, George J., Jr.: Human Transfer Functions in Multi-Axis and Multi-Loop Control Systems. NASA TN D-3305, 1966.
7. Bergeron, Hugh P.; Adams, James J.; and Hurt, George J., Jr.: The Analysis of Human Response in Combined Control Tasks. NASA TN D-4356, 1968.
8. Bergeron, Hugh P.; Kincaid, Joseph K.; and Adams, James J.: Measured Human Transfer Functions in Simulated Single-Degree-of-Freedom Nonlinear Control Systems. NASA TN D-2596, 1965.
9. Adams, James J.; Kincaid, Joseph K.; and Bergeron, Hugh P.: Determination of Critical Tracking Tasks for a Human Pilot. NASA TN D-3242, 1966.

II. QUASI-LINEAR DESCRIBING FUNCTION APPLICATIONS

6. Application of a Systems Analysis Theory for Manual Control Displays to Aircraft Instrument Landing*

*W. F. Clement, H. R. Jex, and D. Graham
Systems Technology, Inc.*

This paper presents an illustrative application of a theory for manual control displays to the instrument landing approach of a large subsonic jet transport. A methodical procedure is disclosed for formulation of compensatory display/control systems.

Manual approach height and lateral position control are treated in multiloop dynamic analyses so as to select preferred variables for measurement and display. Closed-loop system performance and pilot scanning and workload measures are also evaluated.

The example concludes with the prediction of a preferred display arrangement. Comparison with an FAA category II instrument panel arrangement selected by an airline operating the example aircraft shows the predictions to be remarkably accurate.

INTRODUCTION

A comprehensive structure for analyzing manual control display systems has been introduced recently and is currently under development (refs. 1 and 2). It is predicated on the feedback laws governing closed-loop operation of a display/pilot/vehicle system and on simplified but quantitative models for the complex subtasks associated with display utilization. Most of these elements had been separately developed previously by a number of workers, but only recently has it been possible to combine them into a usefully predictive tool for display design. The ingredients of this theory, which is still being developed, are:

System definition.—In systems analysis terms (i.e., mission-based time- and frequency-domain representations of commands, disturbances, merit criteria, and system responses).

Loop topology.—Mathematical empirical models for the elements in the closed-loop task, such as: pilot dynamic response, multiloop vehicular dynamics, display scanning effects, alternative feedbacks available on the displays, control/display associations, and inner-loop/outer-loop display associations.

*This research was accomplished under Contract N00014-66-C0072 for the Office of Naval Research.

Loop closures.—Iterative optimization of the multiloop closures to satisfy various criteria on: stability, performance, and workload.

Closed-loop and displayed signal properties.—Description of the final closed-loop system in terms of its frequency bandwidth, dominant modes, effective delays, scaling, and resolution.

System performance.—Error spectra and relative contributions due to commands, disturbances, and pilot remnant.

Pilot workload.—Metrics for the pilot's workload in scanning, equalizing, and optimizing the task.

Display utilization.—Models and indices for the pilot's use of the display, including fixation traffic among key instruments, and rules and means for improving the arrangement of various displays.

To illustrate the application of this theory to a practical problem, we will use it to design a display system for manually controlled approaches on instrument flight rules in a jet transport aircraft guided by the standard FAA Instrument Landing System (ILS). The display arrangement resulting from this process will be compared at the conclusion with the format actually adopted by an airline using the transport for FAA category II operation.

A flow diagram of the procedure is given in figure 1 with the successive steps shown at the top of the diagram. Although they are presented in sequence, it must be remembered that there is much interaction and iteration among the stages.

SYSTEM DEFINITION

GUIDANCE AND COMMANDS.—The command structure and instrument complex is defined as the ILS guidance scheme adopted by the FAA in which sensed deviations of the

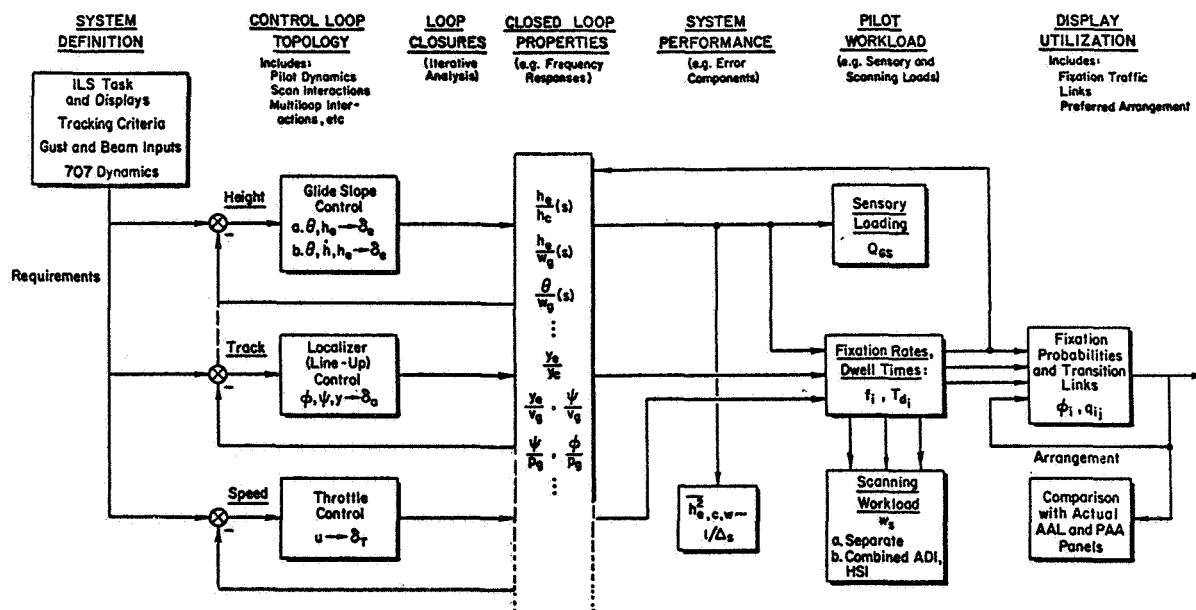


Figure 1.—Flow diagram for the steps in the illustrative example of ILS display for a jet transport.

aircraft from an approximately straight radio beam are displayed to the pilot on a pair of pointers, one for glide slope (height) errors and one for localizer course (lineup) errors. Distortions in the beam null are caused by electrical inhomogeneity of the terrain, yet the aircraft must attempt to follow them in the absence of better position data. Reference 3 presents average power spectral density characteristics for localizer course distortions measured at the receiver output. Corresponding measurements of glide slope receiver output from outer marker to threshold yielded nonstationary mean values, so spectral densities were not computed. Sample time histories of theodolite-corrected glidepath receiver output are presented in reference 3. A typical spectrum of the ILS beam null distortion corresponds to about a 0.5 rad/sec input bandwidth when traversed at typical approach speeds.

DISTURBANCES.—The disturbance structure is assumed to be homogeneous isotropic atmospheric turbulence. Normal and lateral velocity components in airframe body axes have approximate one-dimensional power spectral density given by (refs. 4, 5, and 6):

$$\Phi_{v_g v_g}(\omega) = \Phi_{w_g w_g}(\omega) = \frac{L(\overline{v_g^2} \text{ or } \overline{w_g^2})}{\pi U_o \left[1 + \left(\frac{\omega L}{3\sqrt{U_o}} \right)^2 \right]} \quad (1)$$

where U_o is true airspeed of aircraft, 223 ft/sec

L is integral scale of turbulence, approximately equal to the absolute altitude up to 1000 ft and usually taken as a constant 1000 ft above this altitude

$\overline{v_g^2}$ is mean square lateral gust velocity

$\overline{w_g^2}$ is mean square normal gust velocity

The lateral distribution of normal turbulence velocity produces an effective rolling gust velocity, p_g . Power spectra in reference 7 are low pass and proportional to $\overline{w_g^2} b/L$ (b = wingspan). Half-power bandwidth is approximately inversely proportional to $\sqrt{b/L}$ and directly proportional to U_o/L .

Other disturbance data may be found in reference 8.

TASK CRITERIA.—The basic task will be to control manually the indicated glide slope deviation and localizer errors and indicated airspeed well within acceptable limits, while maintaining adequate safety margins of closed-loop system stability and pilot work load, despite disturbances resulting from gusts and pilot remnant. More quantitatively, these criteria may be defined as:

(1) Error criteria:

Height: $3\sigma_{GSD} \leq A_{GSD} = 50$ ft

Track: $3\sigma_{LOC} \leq A_{LOC} = 100$ ft

Speed: $3\sigma_{IAS} \leq A_{IAS} = 10$ kts

where σ = Root mean square of summed errors due to beam bends, gusts, and pilot remnant
(3σ covers 99.7 percent of all errors)

A = Allowable maximum (rms) task errors

(2) Stability criterion (for the guidance, or outer, loops):

Average gain margin; $KM \geq 3$ dB

Average phase margin; $PM \geq 20^\circ$

(3) Pilot workload criterion:

..... Scanning workload; $W_s \leq 1.0$ (tolerable)

$W_s \leq 0.8$ (desirable)

CONTROLLED ELEMENT DYNAMICS.—The basic vehicle dynamics are those for the Boeing 707-320 in the landing approach configuration, and without yaw damping. The control and gust transfer functions, linearized about the trimmed condition and separated into longitudinal and lateral/directional (three-degree-of-freedom) equations of motion, are based on the data in references 9 and 10.

CONTROL SYSTEM AND NEUROMUSCULAR DYNAMICS.—For illustrative simplicity, the control stick force and displacement characteristics are assumed to be those of a "good" spring-restrained manipulator. A value for the pilot's effective time delay of $\tau_e \doteq 0.3$ second (in the absence of scanning delays) is assumed (ref. 11).

QUASI-CONTINUOUS MODELS.—Because only time-averaged task criteria are to be satisfied, the intrinsically aperiodic nature of the display scanning process can be represented by a quasi-continuous model with an added effective scanning delay, $\Delta\tau$, and sampling remnant injection, n_s (refs. 12 and 13.)

If a combined (integrated) presentation of a few displayed signals having a common control had been anticipated, the results of reference 14 show that the continuous quasi-linear pilot models of reference 11 are valid for describing the adopted multiloop pilot behavior without added scanning delay.

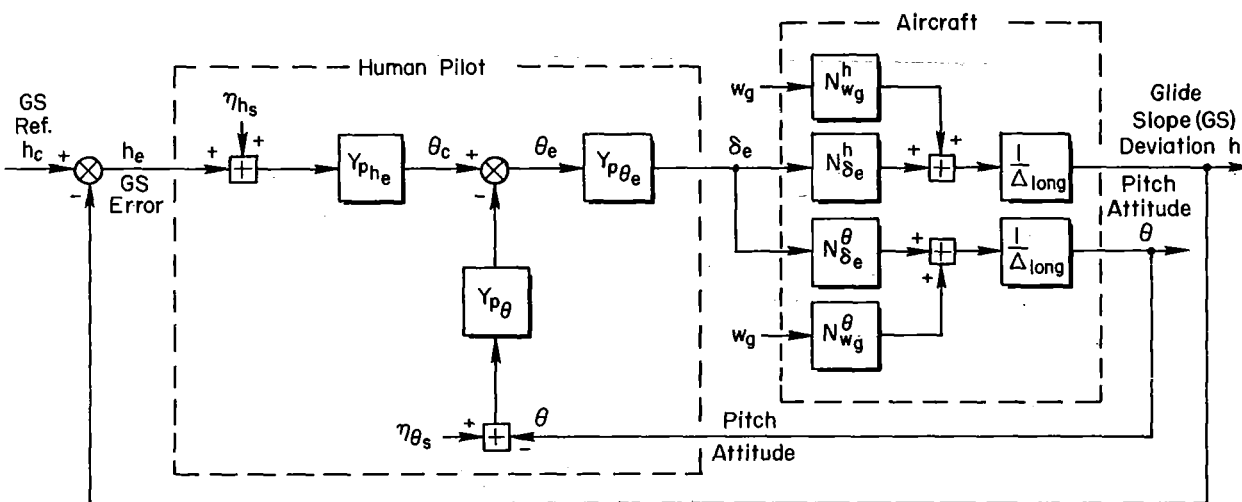
If only two separate displays and control axes were present, the results of reference 15 show also that the continuous quasi-linear pilot models of reference 11 apply to parafoveal as well as foveal perception of the displayed signals without significant scanning delay, albeit with reduced gain.

Because a greater number of separate display presentations are present in instrument landing, some having a common control input and others having separate controls, the relative contribution of parafoveal perception in any one control loop may be so attenuated that an incremental effective scanning delay will exist. Such is the case here. Reference 13 shows, however, that recognition of at least the first time derivative of the displayed signal can negate any scanning delay to a great extent through a process of internal extrapolation and "reconstruction" by the pilot. In either event, the quasi-linear pilot models of reference 11 form the best available basis for describing the adopted multiloop pilot behavior. The time-averaged point of view (in contradistinction to a stimulus-response point of view) greatly facilitates the subsequent analysis, adequately reveals stability problems, and yields valid estimates of the overall closed-loop performance. It is not intended to duplicate the fine grain control behavior.

CONTROL LOOP TOPOLOGY

HEIGHT CONTROL.—Selection of the feedback loops closed by the pilot for the glide slope displacement control task is next performed, using the multiloop feedback selection rules of reference 11. Two alternative display feedback structures are illustrated.

(1) The pilot commands a pitch attitude-to-elevator inner loop based on his perception of glide slope deviation ($h_e \rightarrow \theta_c \rightarrow \delta_e$; fig. 2).



$$Y_{p\theta_e} = e^{-\tau_{\theta_0}s} = e^{-3s}$$

$$Y_{p\theta} = K_{\theta} \left(s + \frac{1}{T_{L\theta}} \right) e^{-\tau_{\theta}s} \\ = -0.63(s+9)e^{-4s}$$

$$Y_{ph_e} = K_{h_e} \left(s + \frac{1}{T_{Lh}} \right) e^{-\tau_{h_e}s} \\ = -0.0013(s+1.1)e^{-4s}$$

$\eta_{h_s}, \eta_{\theta_s}$ are sampling remnant

δ_e elevator control

w_g normal gust velocity

N_a^{β} are airframe transfer function numerators for α input, β output

Δ_{long} is characteristic airframe transfer function denominator for longitudinal dynamics

Approximate Root Loci in Complex Plane

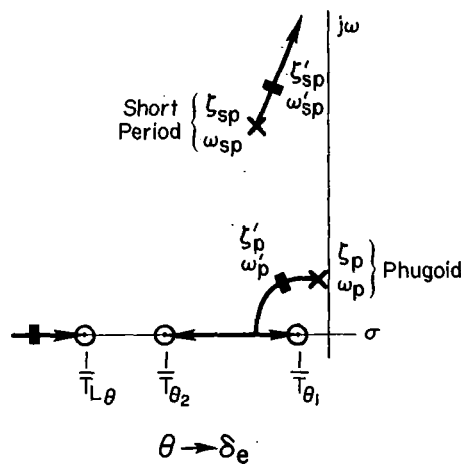
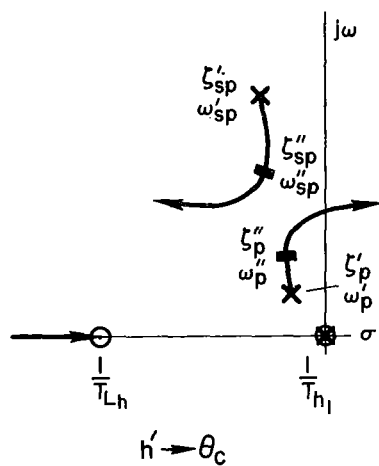


Figure 2.—First alternative: height control system topology for ILS manual control with pitch attitude and glide slope deviation displays.

(2) The pilot commands a pitch attitude inner loop based on his perception of vertical velocity error in a secondary loop with respect to a sink rate reference established by his perception of glide slope deviation ($h_e \rightarrow \dot{h}_c \rightarrow \theta_c \rightarrow \delta_e$; fig. 3). The loops are structured as series rather than parallel loop closures, based on recent results in reference 16.

TRACK (LINEUP) CONTROL. —Selected feedback loops closed by the pilot for the localizer displacement control task are presented in figure 4. The pilot commands a roll attitude to aileron inner loop, based on his perception of heading error in a secondary loop, with respect to a heading reference established by his perception of localizer deviation. ($y_e \rightarrow \psi_c \rightarrow \varphi_c \rightarrow \delta_a$; fig. 4.)

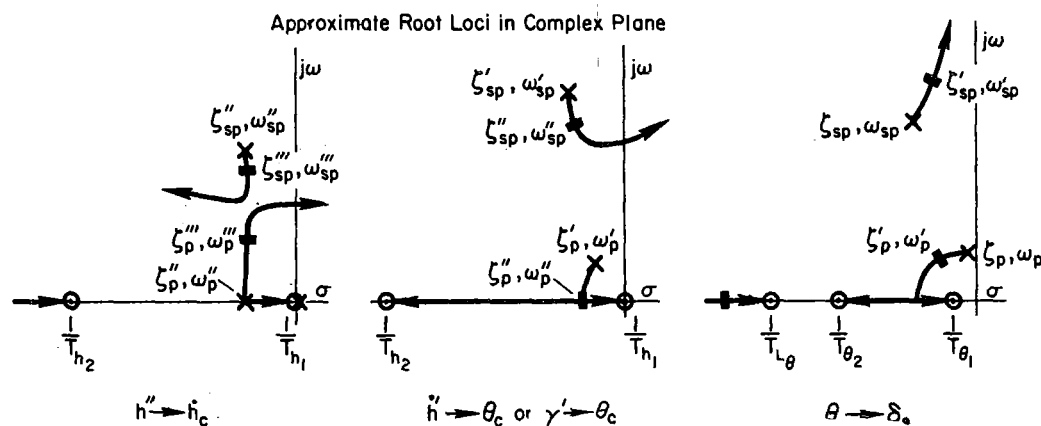
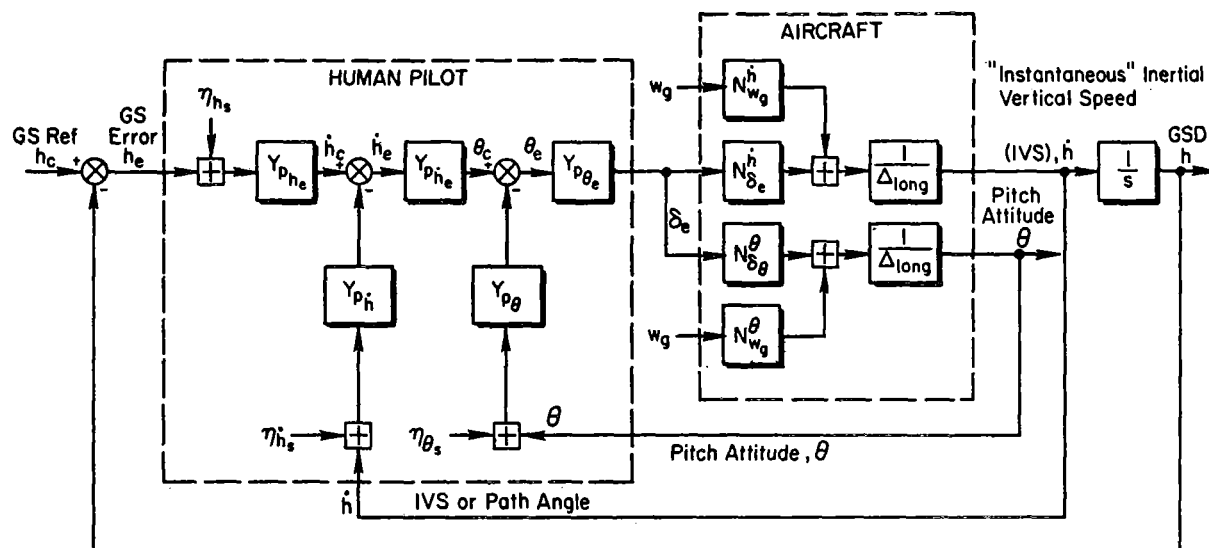


Figure 3. —Second alternative: height control system topology for ILS manual control with instantaneous inertial vertical speed or path angle display in addition to pitch attitude and glide slope deviation.

$$Y_{p\phi_e} = e^{-\tau_{e0}s} \approx e^{-.3s}$$

$$Y_{p\phi} = K_{\phi}(s + 1/T_{L\phi}) = 1.33(s + 1.5)$$

$$Y_{p\psi_e} = K_{\psi_e} \approx 1.0$$

$$Y_{p\psi} = K_{\psi} = 1.9$$

$$Y_{p\psi_e} = K_{\psi_e} = .0018$$

$\eta_{ys}, \eta_{\psi}, \eta_{\phi_s}$ are sampling remnant

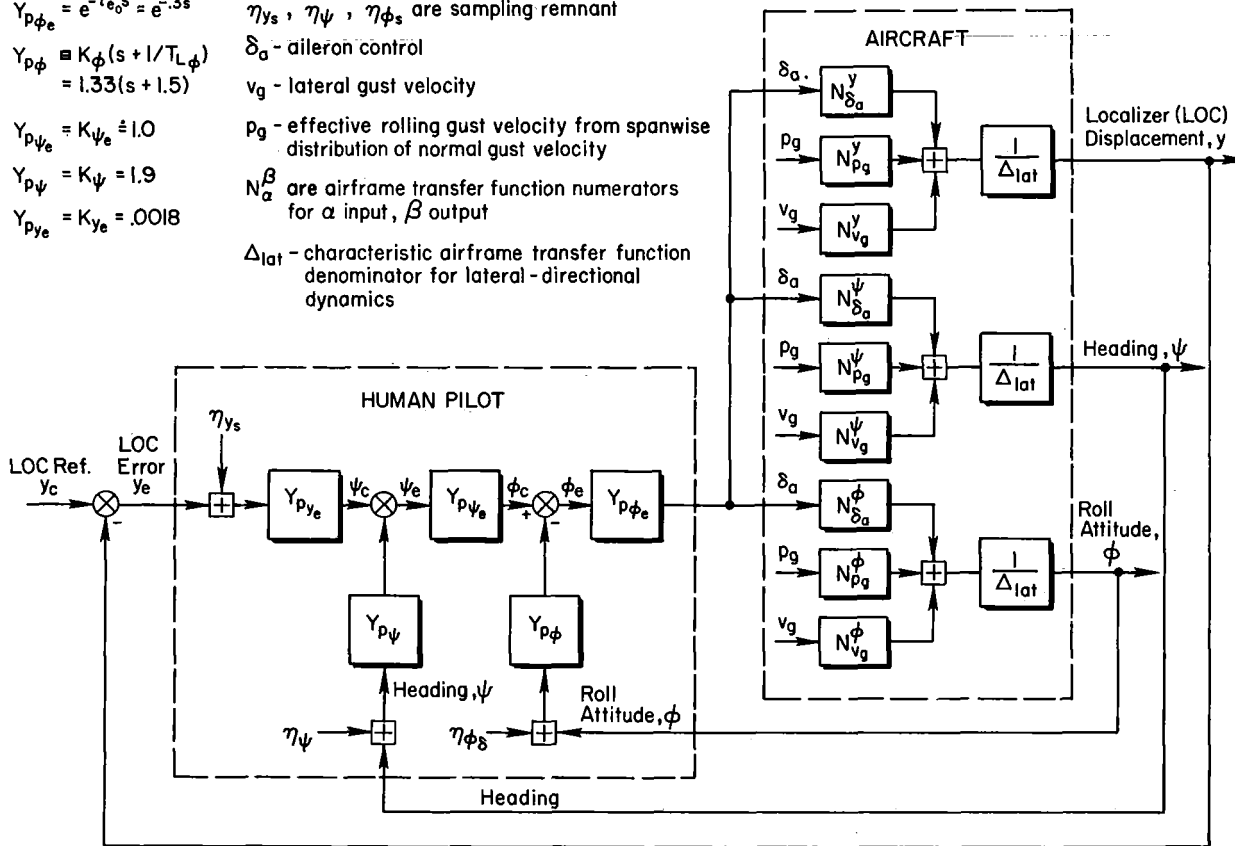
δ_a - aileron control

v_g - lateral gust velocity

p_g - effective rolling gust velocity from spanwise distribution of normal gust velocity

N_{α}^{β} are airframe transfer function numerators for α input, β output

Δ_{lat} - characteristic airframe transfer function denominator for lateral-directional dynamics



Approximate Root Loci in Complex Plane

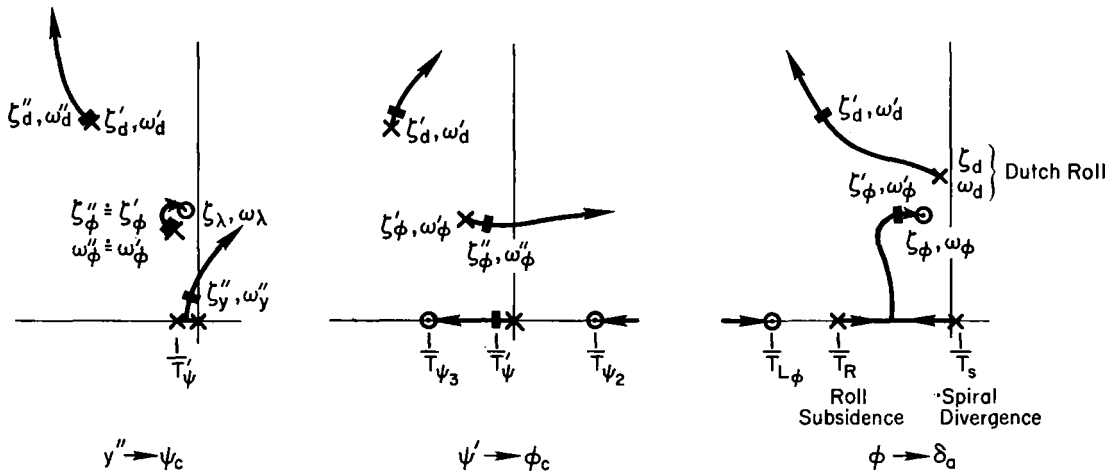


Figure 4. —Track (lineup) control system topology for ILS manual control with roll attitude, heading, and localizer deviation displays.

AIRSPPEED CONTROL. —This loop is not illustrated, but is assumed here to be limited to a simple feedback of the heavily smoothed deviation from the specified indicated airspeed, through the pilot's equalizations to a throttle correction, thence (after considerable thrust lag) to the aircraft speed response. The effects of u-gusts are neglected here, although they should be included with the other gust effects in a complete study of the problem. This loop is further discussed in reference 17.

LOOP CLOSURES

GENERAL. —The complex, multiloop closure process is next performed iteratively until a satisfactory compromise is achieved between the different task criteria of performance, stability, and pilot workload. There is, as yet, no simple paradigm for performing these analyses, because the adapting procedure is, in itself, a multiloop, highly interacting, and nonlinear process that defies any simple analytic or algorithmic prescription. Nevertheless, just as a stable system reaches a set of equilibrium states, so does the pilot's adaptive, scanning, sampling, and equalizing process seem to reach an equilibrium set of parameters (ref. 18). Given an initial loop topology estimate, and some basic loop-closing experience, the analyst can rapidly establish the probable parameters.

Experience has shown that a useful variable during the loop-closing process is the effective scanning delay, $\Delta\tau$, and this is the abscissa in some of the plots to follow.

GLIDE SLOPE DISPLACEMENT LOOP CLOSURE. —The pilot dynamic characteristics have been determined for glide slope displacement control system alternative (a) and are listed in figure 2 for an optimized value of incremental effective scanning delay in addition to the basic effective neuromuscular time delay. Equal values of incremental scanning delay for both pitch attitude and glide slope deviation are used deliberately with some foreknowledge that a common effective bandwidth for both displayed signals will result, because of the rather low frequency short period mode. Equal values of incremental scanning delay simplify the closure analysis without unduly restricting generality in this case, as closed-loop results will subsequently verify. Figure 5 shows how the loop crossover frequencies, displayed bandwidth frequencies, characteristic natural frequencies, and pilot phase lead compensation dynamics vary with total effective pilot time delay, based on the analyses in reference 10.

The four values of incremental effective scanning delay employed in the analysis in figure 5 encompass a range of pilot attention from full to partial attention. The closed-loop short-period frequency (which matches the pitch attitude loop crossover frequency with both loops closed) determines the effective displayed signal bandwidth of pitch attitude and glide slope error.

Next it is necessary to determine the best tradeoffs among the dwell time, scanning delay, and derivative weighting coefficient, λ , in the generalized sampling model developed in reference 13, which will satisfy the bandwidth.

A simultaneous solution for the functional interactions among effective pilot time delay, incremental effective scanning delay, closed-loop displayed signal bandwidth, fixation frequency, dwell time and derivative weighting coefficient is required next. The results of reference 13 show that incremental effective scanning delay can be reduced by one or more stratagems:

- (1) Increasing fixation frequency ($\omega_s = 2\pi f_s = 2\pi/T_s$)
- (2) Increasing derivative weighting factor ($0 \leq \lambda \leq 1$)
- (3) Increasing fixation dwell fraction (T_d/T_s)

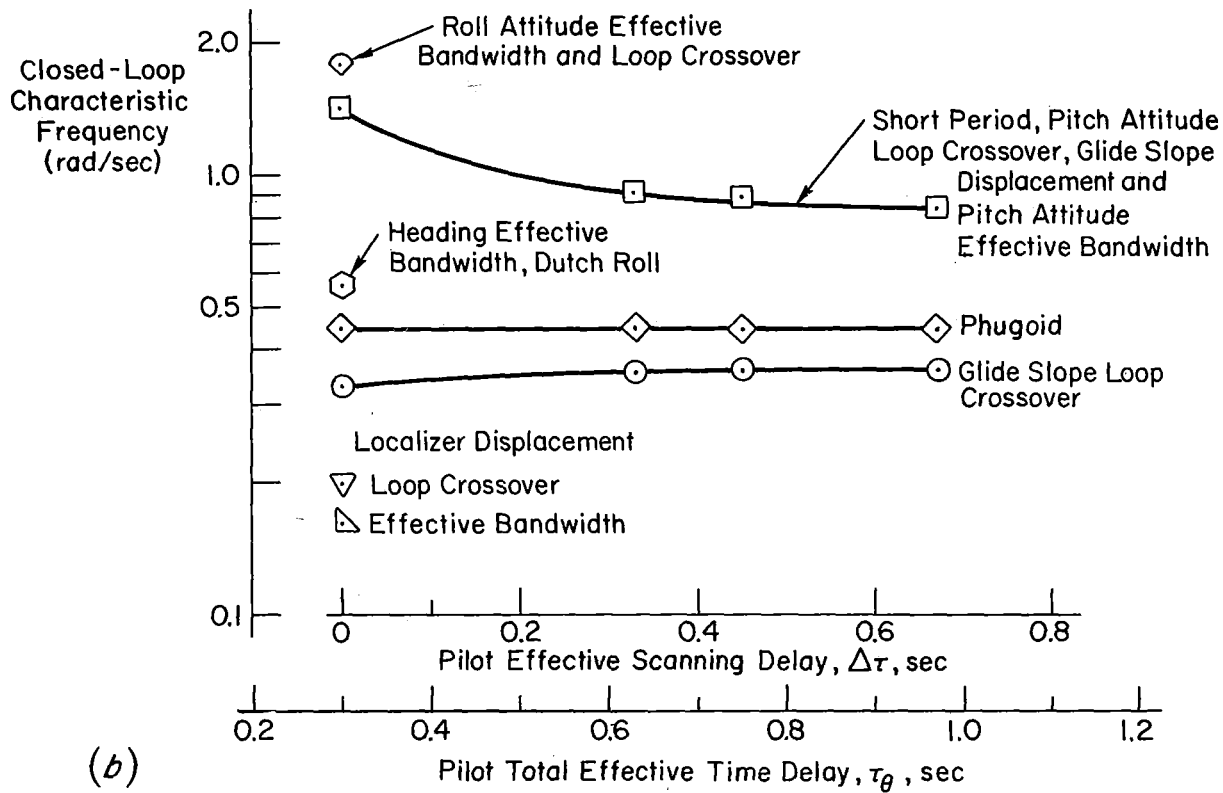
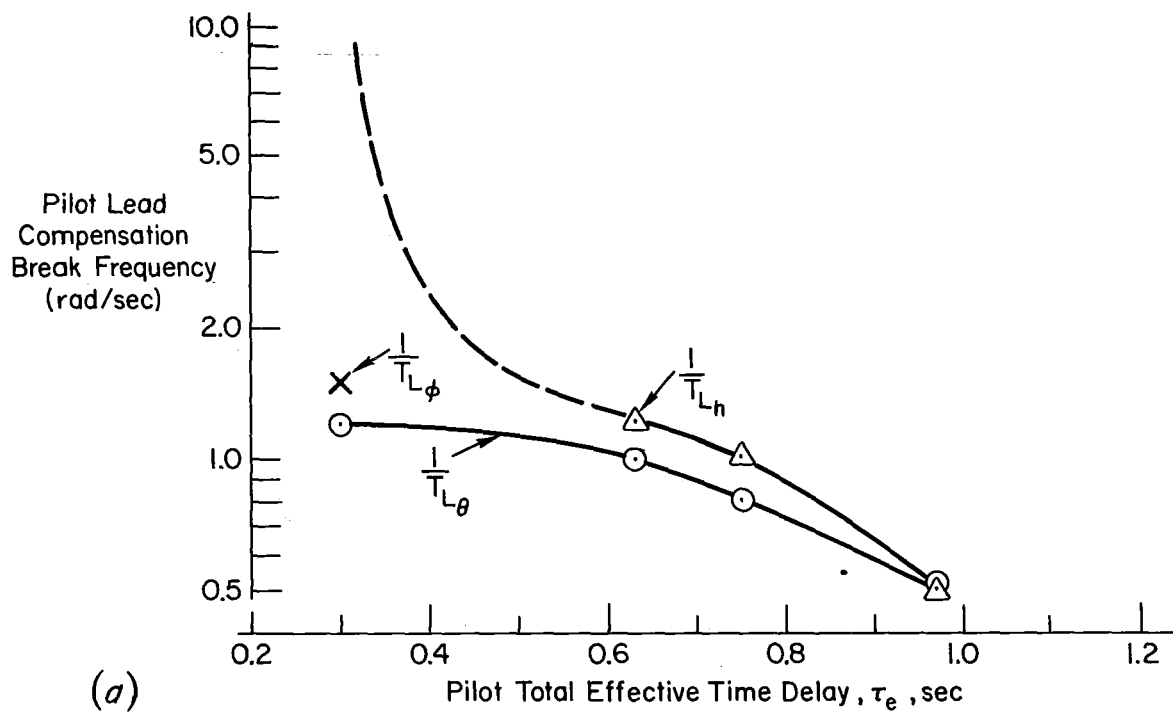


Figure 5.—Characteristic closed-loop frequencies for manual control; B707-320 ILS approach. (a) Pilot equalization. (b) Bandwidth.

A graphic overlay of the results of reference 13 on figure 5 will enable visual inspection of the functional interactions. The combination is shown in figure 6 wherein the overlay is denoted by the dashed curves. This format of overlay was chosen to permit satisfaction of two criteria by merely sliding the overlay:

(1) The total time delay along the abscissa is set by sliding the overlay horizontally to the linear sum of the basic time delay of 0.3 second and $\Delta\tau$ due to scanning.

(2) The ratio between the required fixation frequency, ω_s (on the right log ordinate), and the closed-loop bandwidth, ω_b (on the left log ordinate), is set by sliding the overlay vertically to satisfy the fixation-frequency/bandwidth hypothesis, in this case, $\bar{\omega}_s/\omega_b = 4$. The independent dwell time hypothesis (ref. 10) selects $\bar{T}_d \doteq 0.6$ second and derivative weighting coefficient, $\lambda \doteq 0+$, since extrapolation for signal reconstruction and pilot phase lead compensation (cf. $1/T_{L\theta}$, $1/T_{Lh}$) is required only above $\omega_b = 1$ rad/sec. The preferred simultaneous solution then gives, for a sampling/bandwidth ratio of 4.0, a closed-loop bandwidth, $\omega_b \doteq 0.9$ rad/sec, and total effective delay, $\tau_e = 0.7$ second, denoted by "X." (An alternative solution with scanning frequency reduced to only three times the display bandwidth requires a derivative weighting of $\lambda \geq 0.5$, which is unnecessarily large compared with $1/T_\theta$ and $1/T_h$. On the other hand, an increase in scanning frequency to $\bar{\omega}_s = 5\omega_b$ increases bandwidth imperceptibly, despite a 25 percent increase in fixation frequency. Therefore, choices other than X yield diminishing returns and will cause unnecessary scanning work load.)

This overlay solution allows changes in assumptions or pilot behavior to be readily seen and illustrates one advantage of this analytic approach.

The pilot dynamic characteristics will be similar for the alternative glide slope displacement display control system (b). They are not discussed in detail. Figure 3 illustrated the trends in the system root locus as the three loops, $\theta \rightarrow \delta_e$, $h \rightarrow \theta_c$, $h \rightarrow \dot{h}_c$, were successively closed. Display control system (b) will result in closed-loop bandwidth approximately equivalent to that for system (a) without requiring pilot phase lead compensation, per se, in the glide slope deviation loop ($1/T_{Lh} \rightarrow \infty$). Contrarily, scanning workload may increase, because perception of another signal, "instantaneous" vertical velocity (or the dynamically equivalent vertical flight path angle) and closure of its loop are required in lieu of phase lead compensation on h . This increment in scanning workload need not arise if the vertical velocity (or path angle) display can be sufficiently integrated with the deviation or pitch angle display as to minimize scanning penalties.

LOCALIZER DISPLACEMENT LOOP CLOSURE.—The pilot dynamic characteristics have been determined in reference 10 for localizer displacement control and were listed in figure 4 for the limiting case of full attention to all three signals, i.e., no incremental scanning delay in addition to the basic effective reaction time delay, $\tau_e = 0.3$ second. Resulting closed-loop characteristic frequencies and effective displayed signal bandwidth frequencies were shown in figures 5 and 6. Localizer bandwidth is so low, $\omega_b = 0.2$ rad/sec, that either very low frequency phase lead compensation, or very large first-derivative weighting ($1/T_{Ly} < 0.2$ or $\lambda > 0.97$, if $\bar{\omega}_s = 4\omega_b$), or very high fixation frequency in relation to bandwidth ($\bar{\omega}_s \doteq 12\omega_b$) is required to compensate for incremental scanning delay. Integration of localizer deviation display with heading or roll attitude and reliance on common fixation or parafoveal perception is one solution which will increase fixation frequency without scanning work load. Incorporation of low frequency compensation in a localizer flight director display is another possibility. Automatic localizer displacement control in a split-axis arrangement is yet another possibility for relieving low frequency reconstruction and scanning work load. In this example, we shall eventually adopt the first alternative—combined display and relatively high scanning frequency.

Hypothesis $\bar{\omega}_s = 4\omega_b$
 Basic Reaction Time Delay .3 sec
 for Full Attention

Mean Ocular Fixation
 Frequency, $\bar{\omega}_s$ (rad/sec)

LOOPS
 Roll \diamond
 Pitch and GSD \square
 Heading \odot
 Speed \diamond
 Localizer ∇

Preferred Solution

$\frac{\bar{\omega}_s}{\omega_b} = 4$

0
 .5
 .75
 .90
 .97

First
 Derivative
 Weighting, λ
 (Dwell time 0.6 sec)

(Scale for solid
 curves)

Pilot Incremental Effective Scanning Delay, $\Delta\tau$, sec

$\tau_{e0} = .3$ sec

(Scale for dashed
 curves)

Pilot Total Effective Time Delay, τ_e , sec

79

The effective heading bandwidth ($\omega_b \doteq 0.6$ rad/sec) is also sufficiently low that large first-derivative weighting ($\lambda \doteq 0.9$) is required to compensate for incremental scanning delay if $\bar{\omega}_s = 4\omega_b$. This result implies that a lateral flight director display (equalized error display) would provide a significant improvement in performance over the ILS scheme.

The effective roll attitude bandwidth is sufficiently high that only modest first-derivative weighting ($\lambda \doteq 0.5$) or modest decrease in pilot lead ($1/T_{L\phi}$ to 1 rad/sec) is required to compensate for incremental scanning delay if the scanning frequency is still four times the bandwidth.

AIRSPEED LOOP CLOSURE.—Although analysis of an airspeed-to-throttle control loop has been omitted in this paper, the bandwidth of displayed airspeed will correspond to that of the closed-loop phugoid and is shown in figure 6. A derivative-weighting coefficient, $\lambda \doteq 0.8$ if $\bar{\omega}_s \doteq 3\omega_b$, is estimated to satisfy airspeed scanning requirements, since power-plant delay and lag usually dominate the airspeed-to-throttle loop. Manual airspeed control via throttle is further discussed in reference 17.

The analysis of human pilot perception and control dynamics for this ILS approach control problem has now reached the point where closed-loop dynamics and average performance can be derived prior to finding measures of workload and display arrangement.

CLOSED-LOOP PROPERTIES

HEIGHT CONTROL.—The height control loop has a fifth-order characteristic equation with adequate stability margins. The closed-loop frequency responses of glide slope displacement are shown in figures 7 and 8. Each figure presents the amplitude ratio of the frequency response to a particular forcing function and the normalized power spectral density of the forcing function on logarithmic frequency scales. A Cartesian graph then illustrates the normalized power spectral density of the response obtained by multiplying the squared modulus of the response by the power spectral density of the command or gust. Figure 7 gives the error response to beam null distortion, which is desirably low over the path control frequency range (up to 0.3 rad/sec) and equal to unity (negligible response) above 1.0 rad/sec. However, in the range of 0.5 to 1.0 rad/sec there is an error response peak that amplifies beam noise in this range. This provides the dominant frequencies seen on the GSD needle which the pilot must track, and justifies the previous computations of $\omega_b \doteq 0.9$ rad/sec.

Figure 8a gives the height deviation response to vertical gusts, a typical spectrum for which is shown in figure 8b corresponding to a 700-foot scale wavelength of turbulence. According to figure 8c the displayed signal bandwidth due to gusts is also about 1.0 rad/sec.

The closed-loop response of pitch attitude to normal gust velocity is shown in figure 9. Again the displayed signal bandwidth is 1.0 rad/sec.

TRACK (LINEUP) CONTROL.—Figure 10 presents localizer displacement error response to radio beam null irregularities. The peak amplification of localizer null distortion will occur in the frequency range from 0.2 to 0.5 rad/sec. A typical waveguide localizer noise spectrum is illustrated. The effective bandwidth of localizer displacement is 0.2 rad/sec. Notice the example of localizer crossover frequency regression, because of the stability limitation. Error suppression bandwidth desirably underlaps command bandwidth, because the pilot does not want to follow beam irregularities anyway and applies some spectral discrimination in practice to attempt to reject command noise from significant error.

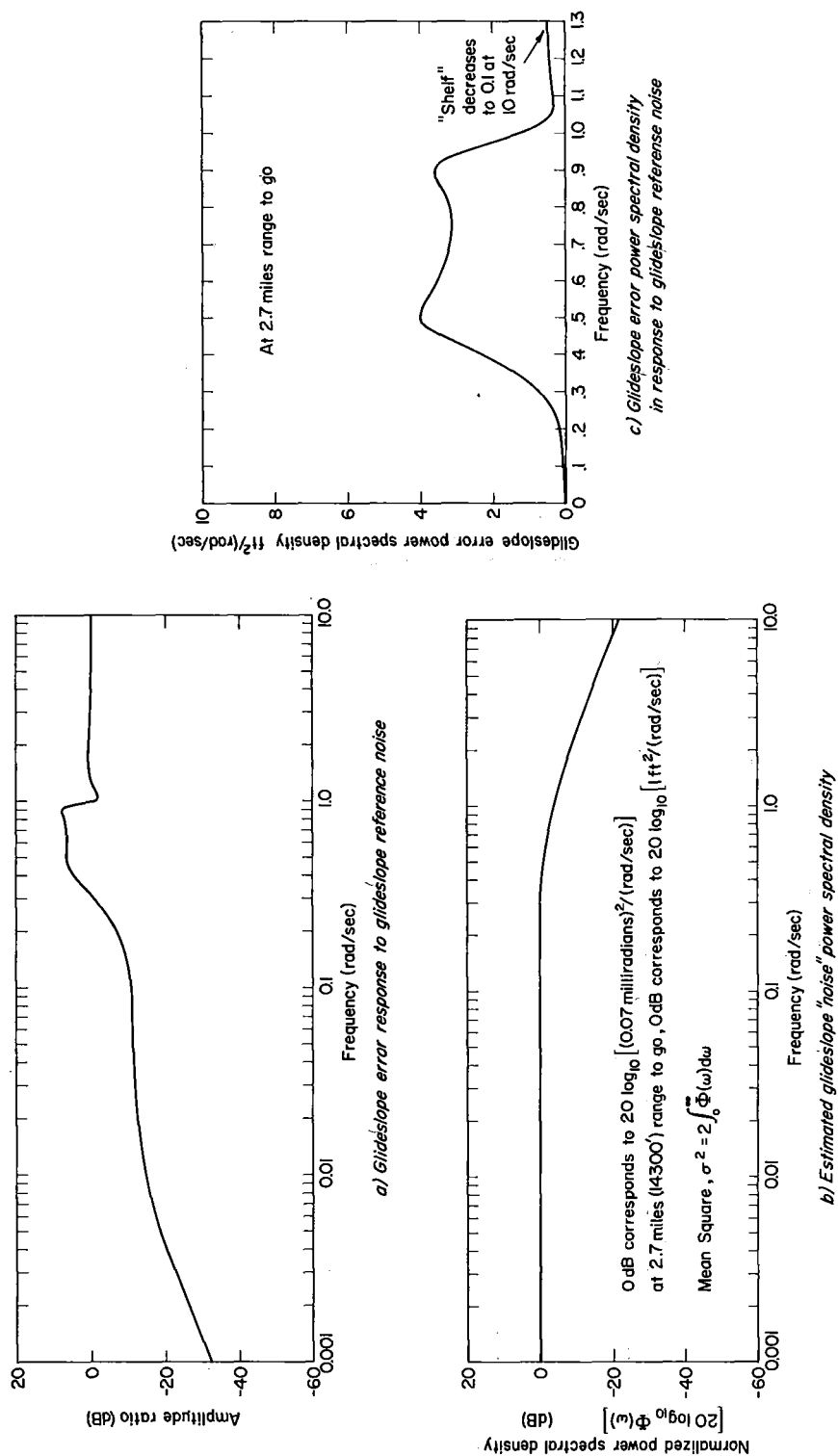


Figure 7.—Closed-loop glide slope error and glide slope noise spectral characteristics.

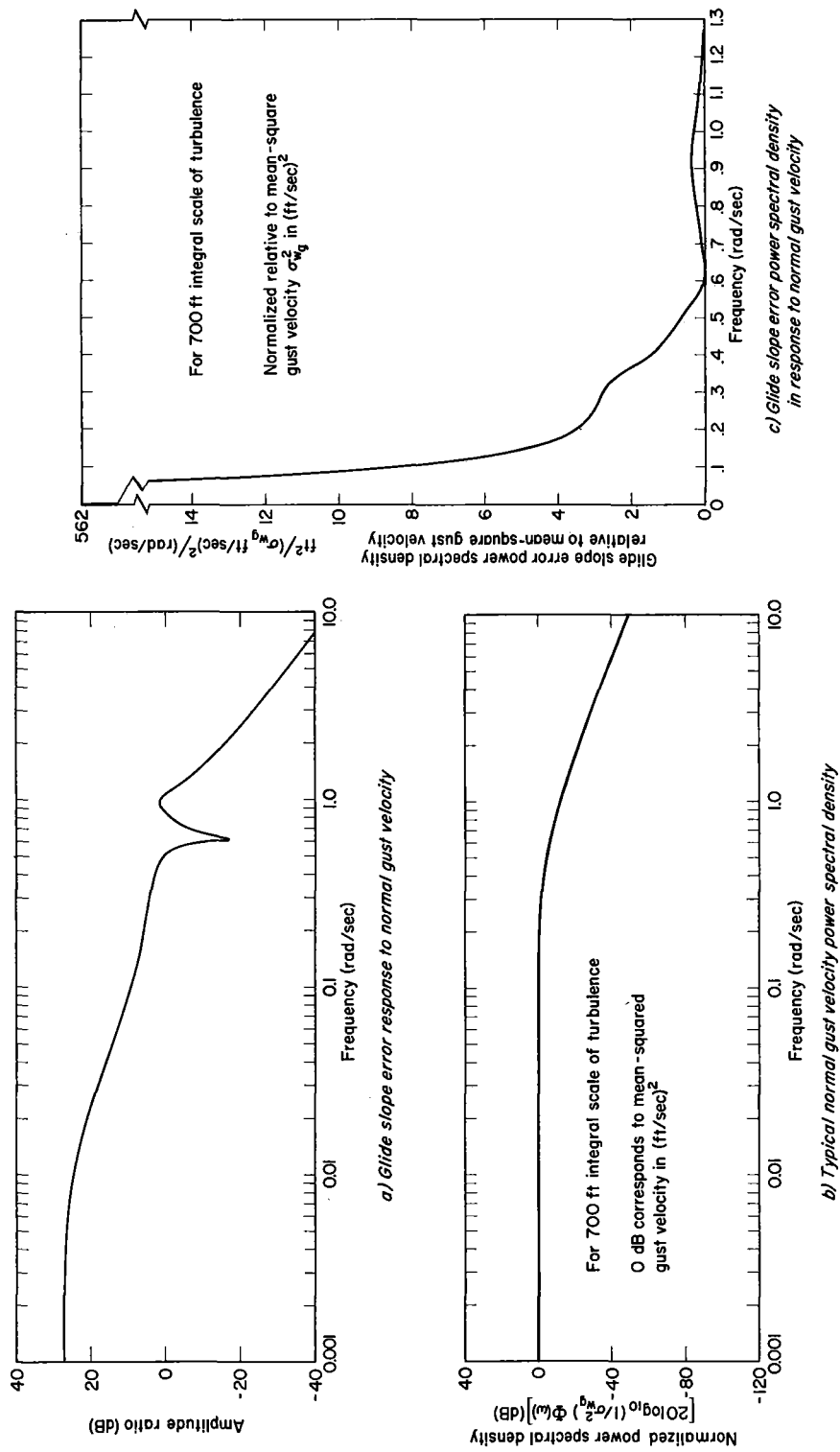


Figure 8.—Closed-loop glide slope error and normal gust velocity spectral characteristics.

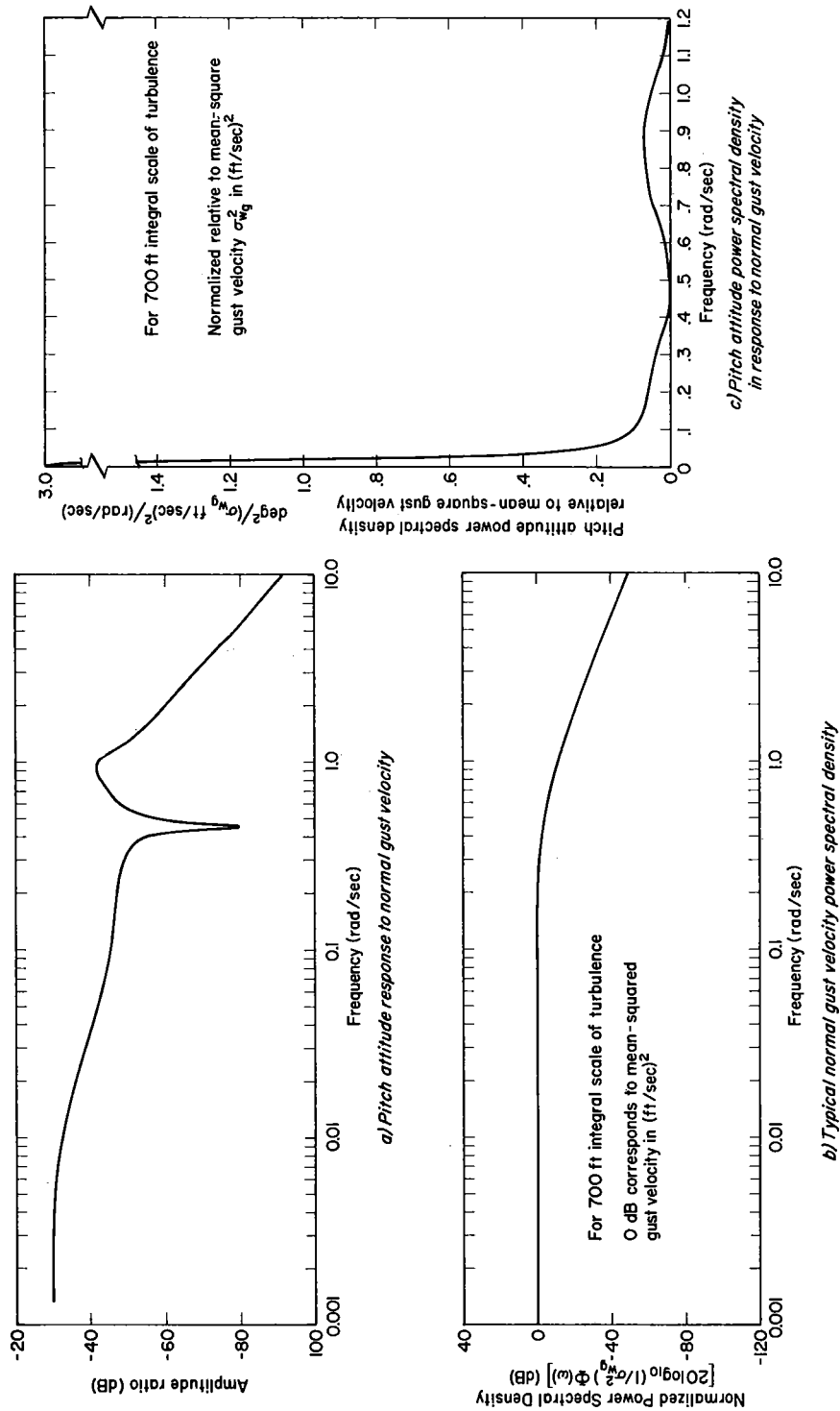


Figure 9.—Closed-loop pitch attitude and normal gust velocity spectral characteristics.

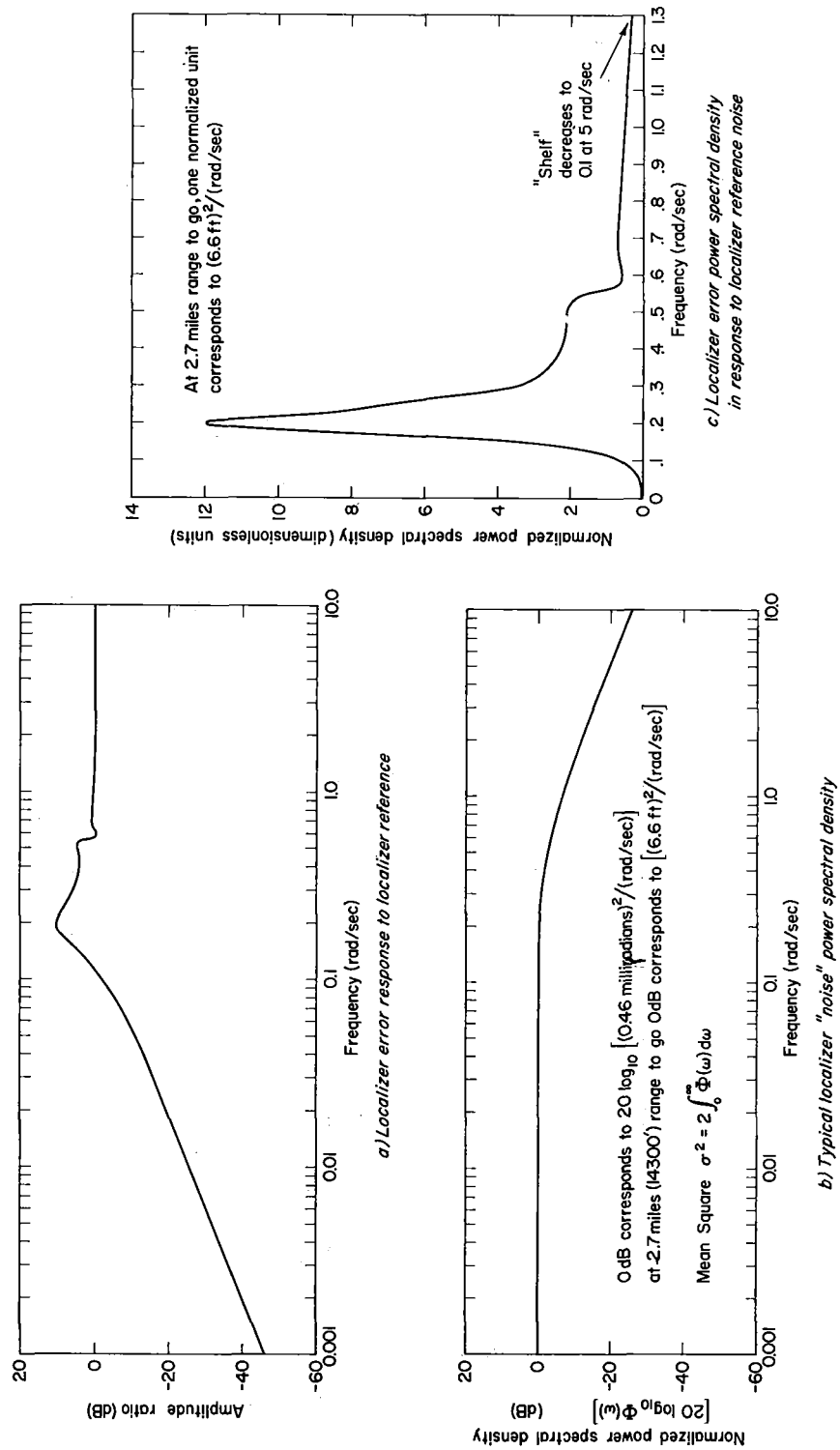


Figure 10. —Closed-loop localizer error and localizer noise spectral characteristics.

In like fashion localizer displacement response to lateral gust velocity can be shown to have a displayed bandwidth of about 0.2 rad/sec (ref. 10).

Heading frequency responses to lateral and effective "rolling" gust velocities are determined in reference 10. The displayed bandwidth is dominated by the closed-loop dutch roll frequency, 0.6 rad/sec in both cases.

Finally, roll attitude response to "rolling" gust velocity is determined in reference 10. The displayed signal bandwidth is about 2 rad/sec.

SYSTEM PERFORMANCE MEASURES

The ILS glide slope tracking performance of the complete display/pilot/aircraft system is now computed to illustrate the procedures involved.

Relationships for sum-squared glide slope error and pitch attitude are derived in reference 10 for the loop topology of figure 3 based on a random model for the sampling remnant (ref. 12). The method can be extended to include three or more loops. In this formulation, the circulating errors in the multiloop system introduced by random sampling having an average scan interval of \bar{T}_s seconds are accounted for simply by multiplying the continuous nonsampled mean-squared errors by a factor of $1/\Delta_s$, greater than unity. (The continuous model contains the effective sampling delays from cardinal reconstruction theory in reference 13; Δ_s is a sampling error determinant.)

In practice the coupling products in reference 10 involving contributions from mean-squared pitch attitude are small. If these contributions are neglected, and the crossover model (ref. 11) of the resulting glide slope loop is used, an estimate of relative mean-squared glide slope error contributions from command reference noise and normal turbulence can be made rather easily.

The sum-squared glide slope deviation, S_{GSD}^2 , as derived in reference 10 is given approximately by:

$$S_{GSD}^2 = \frac{1}{\Delta_s} \left[\left(\frac{\bar{h}_e^2}{\bar{h}_c^2} \right)_{w_g=0} \bar{h}_c^2 + \left(\frac{\bar{h}_e^2}{\bar{w}_g^2} \right)_{h_c=0} \bar{w}_g^2 \right] \quad (1)$$

where the glide slope command-following error is given by:

$$\left(\frac{\bar{h}_e^2}{\bar{h}_c^2} \right)_{w_g=0} = \frac{1}{1 - 2\omega_c \tau_e} \left[1 - \frac{\omega_c / \omega_{hc}}{\sqrt{1 - 2\omega_c \tau_e}} \arctan \left(\frac{\omega_{hc}}{\omega_c} \sqrt{1 - 2\omega_c \tau_e} \right) \right] \quad (2)$$

provided $\omega_{hc} \tau_e < 1.0$

This is based on the crossover model of reference 11 and a quasi-rectangular input spectrum having bandwidth ω_{hc} and mean square \bar{h}_c^2 (ref. 23). For $\omega_{hc} \doteq 1$ rad/sec, $\omega_c \doteq 0.36$ rad/sec and $\tau_e = 0.7$ second from the closed-loop characteristics and pilot dynamics,

$$\left(\frac{\bar{h}_g^2}{\bar{h}_c^2}\right)_{w_g=0} \doteq 0.86 \quad (3)$$

The next term required in equation (1) is given by

$$\left(\frac{\bar{h}_g^2}{\bar{w}_g^2}\right)_{h_c=0} \doteq \frac{(\bar{L}/\pi U_0)}{\omega_{wg} \omega_c \sqrt{1-2\omega_c \tau_e}} \arctan \left(\frac{\omega_{wg}}{\omega_c} \sqrt{1-2\omega_c \tau_e} \right)$$

provided $\omega_{wg} \tau_e \ll 1.0$ (4)

This is derived in reference 10 from the crossover model for vertical gust velocity inputs. For $L = 700$ feet, $\omega_{wg} = 0.77$ rad/sec and

$$\left(\frac{\bar{h}_g^2}{\bar{w}_g^2}\right)_{h_c=0} \doteq \frac{700}{223\pi} \left(\frac{14}{3}\right) \doteq 4.7 \quad (5)$$

The next factor required in equation (1) is

$$\Delta_s \doteq 1 - \frac{\bar{T}_{sh_e}}{\pi} \int_0^\infty \left| \frac{h_e}{n_{hs}} \right|^2 d\omega \quad (6)$$

and derived from the crossover model for random sampling remnant in reference 10.

For

$$\bar{T}_{sh_e} = \frac{2\pi}{\omega_{sh_e}} = \frac{2\pi}{4} \text{ sec}$$

$$\Delta_s \doteq \frac{2}{3} \quad (7)$$

When the numerical results of equations (3), (5), and (7) are substituted into equation (1), the sum-squared glide slope deviation becomes

$$s_{GSD}^2 \doteq 1.5 \left[0.86 \bar{h}_c^2 + 4.7 \bar{w}_g^2 \right] \quad (1a)$$

With a particularly noisy beam (e.g., Burbank) the rms glide slope beam deviation is on the order of $\sigma_{\text{beam}} \doteq 0.05^\circ \doteq 0.001 \text{ rad}$ (ref. 3), and at about a 10,000-foot range this corresponds to an rms altitude command of $\sigma_{h_c} \doteq 0.001 \times 10,000$. A corresponding value for the rms gust velocity in rough air at low altitudes is about $\sigma_{w_g} = 3 \text{ ft/sec}$. Typical closed-loop glide slope tracking errors are then given by:

$$\begin{aligned} S_{\text{GSD}}^2 &\doteq 1.5[0.86(10)^2 + 4.7(3)^2] \\ &\doteq 1.5[86 + 42] = 192 \text{ ft}^2 \end{aligned} \tag{1b}$$

$$S_{\text{GSD}} \equiv \sigma_{h_e} \doteq 14 \text{ ft}$$

$$\text{Also: } 3\sigma_{h_e} \doteq 42 \text{ ft} < A_{h_e} = 50 \text{ ft}$$

Inspection of this last equation reveals that:

- (1) The glide slope tracking errors due to beam noise are greater than those due to turbulence.
- (2) The sampling remnant increases the errors by 50 percent over the continuous system values.
- (3) The total 3σ error is within the specified limits.

The procedure for estimating sum-squared localizer deviation is analogous.

Once these analytical/numerical expressions for the system performance are available, it becomes very easy to evaluate its sensitivity to the display system design variables and pilot utilization parameters. Concepts and refinements that result in only a small fractional improvement can be weeded out and the really helpful or disastrous changes are clearly revealed prior to undertaking costly simulator experiments.

Because the system has been represented by quasi-linear models, all the signals in the loop can be obtained by appropriate ratios such as those in equation (1), and the levels of control activity, display signals, and accelerations on the pilot can easily be assessed by this approach.

SCANNING WORKLOAD

The scanning workload, W_s , measures the fraction of the pilot's available time spent scanning an individual or a group of displays (refs. 2, 10, and 20). Abbreviations for displays are defined in table 1. The maximum possible value of W_s is unity. The first trial calculation of scanning workload is presented in table 2, wherein each signal required except roll and pitch attitude (combined on an artificial horizon) is assigned a distinct display and corresponding fixation. The mean fixation frequency hypothesis and display/control frequency bandwidth have already been discussed for the key selected loops in the system topology of figure 3 and figure 4. Vertical speed (conventional lagged aneroid, not instantaneous), pressure and radio (absolute) altitude, and turn and bank (needle and ball) do not participate in primary display/control loops; therefore, they are accorded lower mean fixation frequency hypotheses and "monitor" mean dwell times, as explained in reference 10. Localizer displacement and heading are assigned slightly longer mean dwell times than glide slope displacement and attitude because of the large derivative weighting coefficients required by the former pair. The resulting scanning workload, $W_s = 1.32 > 1$, exceeds the pilot's limits; therefore, integrated displays, equalized displays ("quickenings"),

TABLE 1. —DISPLAYED SIGNAL AND INSTRUMENT ABBREVIATIONS

ADI	Attitude director	IVS	Instantaneous inertial vertical speed
AS	Airspeed		
		LOC	Localizer deviation
ATT	Artificial horizon (roll and pitch attitude)	MISC	Miscellaneous
ENG	Engine	PALT	Pressure altitude
GSD	Glide slope deviation	RALT	Radio altitude
HDG	Heading	RMI	Radio magnetic (bearing) indicator
HSI	Horizontal situation indicator	T and B	Turn and bank (needle and ball)
IAS	Indicated airspeed	VS	Vertical speed

a flight director (where λ is large), split-axis automatic/manual control, or some combination of these is required to reduce scanning workload within the criterion $W_s \leq 1$. A conventional combination of displays will be adopted to illustrate the tradeoff.

Table 3 presents scanning workload for display combinations adopted in accord with those of Pan American World Airways for FAA category II certification. Since $W_s \doteq 1$, scanning workload is just saturated, and this combination of displays is at least practical for this problem.

Tables such as 2 and 3 are especially useful in evolving and evaluating new display concepts because the tradeoffs among the signal bandwidth (which limits the available scanning interval), the number of places to be fixated (which goes down with combined displays), and dwell times required to assimilate the information (which goes up with combined displays) are all clearly revealed in the Fractional scanning workload column. For example, in this case, comparison of Tables 2 and 3 reveals that combining the GSD, LOC, ATT, and T and B presentations on one display increases the required dwell time from 0.6 to 0.8 second, but decreases the proportion of total time spent on these instruments from 86 percent to 51 percent.

DISPLAY UTILIZATION

FIXATION PROBABILITIES. —As defined in references 19 and 20, the probability, p_i , of fixating a given instrument in an ensemble of fixations is estimated by φ_i :

$$p_i \doteq \varphi_i = \frac{\pi_i \bar{T}_{d_i}}{\sum_i \pi_i \bar{T}_{d_i}} \text{ where } \pi_i = \frac{\bar{f}_{s_i}}{\sum_i \bar{f}_{s_i}}$$

TABLE 2.—SCANNING WORKLOAD WITH SEPARATE DISPLAYS

Displayed signal candidate	Fixation frequency hypothesis ω_{s_i}/ω_b	Display/control frequency bandwidth		Mean fixation frequency		Pilot's role and equalization $1/T_L = \text{Phase lead break frequency}$ $\lambda = \text{Derivative weighting coefficient}$	Mean dwell time sec \bar{T}_{d_i}	Fractional scanning workload $W_{s_i} = \bar{f}_{s_i} \bar{T}_{d_i}$
		rad/sec	ω_b	rad/sec	H_z			
				ω_{s_i}	\bar{f}_{s_i}			
GSD	4	1.0		4.0	0.637	$1/T_L = 1.1$ $\lambda = 0$	0.6	0.382
ATT	4	1.0		4.0	.637	$1/T_L = 0.9$ (pitch) $\lambda = 0$ $1/T_L = 1.5$ (roll) $\lambda = 0.5$.6	.382
AS	3	0.4		1.2	.191	$\lambda = 0.8$.5	.096
VS	1	.4		0.4	.0637	Monitor	.4	.025
PALT	1	.4		.4	.0637	Monitor	.4	.025
RALT	1	.4		.4	.0637	Monitor	.4	.025
LOC	4	.2		.8	.127	$\lambda = 1$.7	.089
HDG	4	.6		2.4	.382	$\lambda = 0.9$.7	.267
RMI								
T&B	1/3	.6		.2	.0318	Monitor	.4	.013
ENG		(Estimated)		.1	.0159	Monitor	.8	.013
MISC								

Total scanning workload: $W_s = \sum_i \bar{f}_{s_i} \bar{T}_{d_i} = 1.32 > 1$

∴ Scanning workload is oversaturated

∴ Combined signal displays are required

TABLE 3.—SCANNING WORKLOAD WITH SOME COMBINED DISPLAYS

Displayed signal candidate	Fixation frequency hypothesis $\bar{\omega}_{s_i}/\omega_b$	Display/control frequency bandwidth		Mean fixation frequency		Pilot's role and equalization $1/T_L$ = Phase lead break frequency λ = Derivative weighting coefficient	Mean dwell time	Fractional scanning workload
		rad/sec	ω_b	rad/sec	$\bar{\omega}_{s_i}$			
							\bar{T}_{d_i}	$W_{s_i} = \bar{f}_{s_i} \bar{T}_{d_i}$
GSD LOC ATT T&B	4	1.0	4.0	0.637	$1/T_L = 1.1$ (GSD) $\lambda = 0$ (LOC) $\lambda = 1$ $1/T_L = 0.9$ (pitch) $\lambda = 0$ $1/T_L = 1.5$ (roll) $\lambda = 5$	0.8	0.51	
AS	3	0.4	1.2	.191	$0 < \lambda < 1$.5	.096	
VS	1	.4	0.4	.0637	Monitor	.4	.026	
PALT	1	.4	.4	.0637	Monitor	.4	.026	
RALT	1	.4	.4	.0637	Monitor	.4	.026	
LOC HDG	4	.6	2.4	.382	(LOC) $\lambda = 1$ (HDG) $\lambda = 1$.8	.306	
RMI		(Estimated)	.1	.0159	Monitor	.4	.006	
ENG		(Estimated)	.1	.0159	Monitor	.8	.013	
MISC		(Estimated)	.1	.0159	---	.2	.003	

Total scanning workload: $W_s = \sum_i \bar{f}_{s_i} \bar{T}_{d_i} = 1.01 \div 1$

\therefore Scanning workload is just saturated

φ_i differs from W_{s_i} in that it measures the relative frequency number of fixations on a given instrument, rather than the relative time it is in use.

The resulting fixation probabilities are shown superimposed on an array of instruments in figure 11 for the combination covered in table 3. It is readily apparent that the combined-display instruments (ADI and HSI) receive about 80 percent of the pilot's fixations.

TRANSITION LINKS.—The relative number of fixation transitions between pairs of instruments is measured by the two-way transition link probability (ref. 20):

$$(q_{ij} + q_{ji}) = \frac{2\varphi_i\varphi_j}{1 - \sum_k \varphi_k^2}$$

These have also been computed for the case of table 3 and are shown by the relative width of line between instruments in figure 11.

DISPLAY ARRANGEMENT

Probabilities of fixation and link values are calculated in reference 2 for the system of table 3, and the results appear in the arrangement of figure 11, based on the display arrangement hypothesis (refs. 10, 20, 21, and 24):

- (1) Locate centrally those displays having the highest probability of fixation
- (2) Locate peripherally adjacent those displays having highest link values with the central display(s)
- (3) Locate peripherally remote those displays having lowest probability of fixation and/or lowest link values

The predicted arrangement agrees with that actually adopted for Pan American-FAA category II certification as shown in the inset of figure 11 (ref. 22).

A slightly different arrangement has been determined in reference 10 for the system topology of figure 3 and figure 4. This system employs the instantaneous vertical speed display previously discussed. The predicted arrangement differs slightly from that actually adopted for American Airlines-FAA category II certification (ref. 22). However, differences in location of pressure altitude and radio magnetic bearing are reasonable in view of navigation scanning convenience in flight conditions other than approach.

CONCLUSIONS

This analysis of piloted display utilization during the landing approach of a large jet transport via ILS guidance has shown that the comprehensive display theory of references 1 and 2 can successfully cope with the numerous factors involved such as: various inputs disturbances and criteria; complex multiloop dynamics; pilot's display scanning effects; system performance and pilot workload. In addition, the theory reveals the strong interactions among these factors. The face validity of the numerical results is strengthened by: the particular display feedbacks revealed as best, the high scanning and workload requirements, the reasonable error levels and bandwidths, and similarity of the derived display arrangement to airline practice.

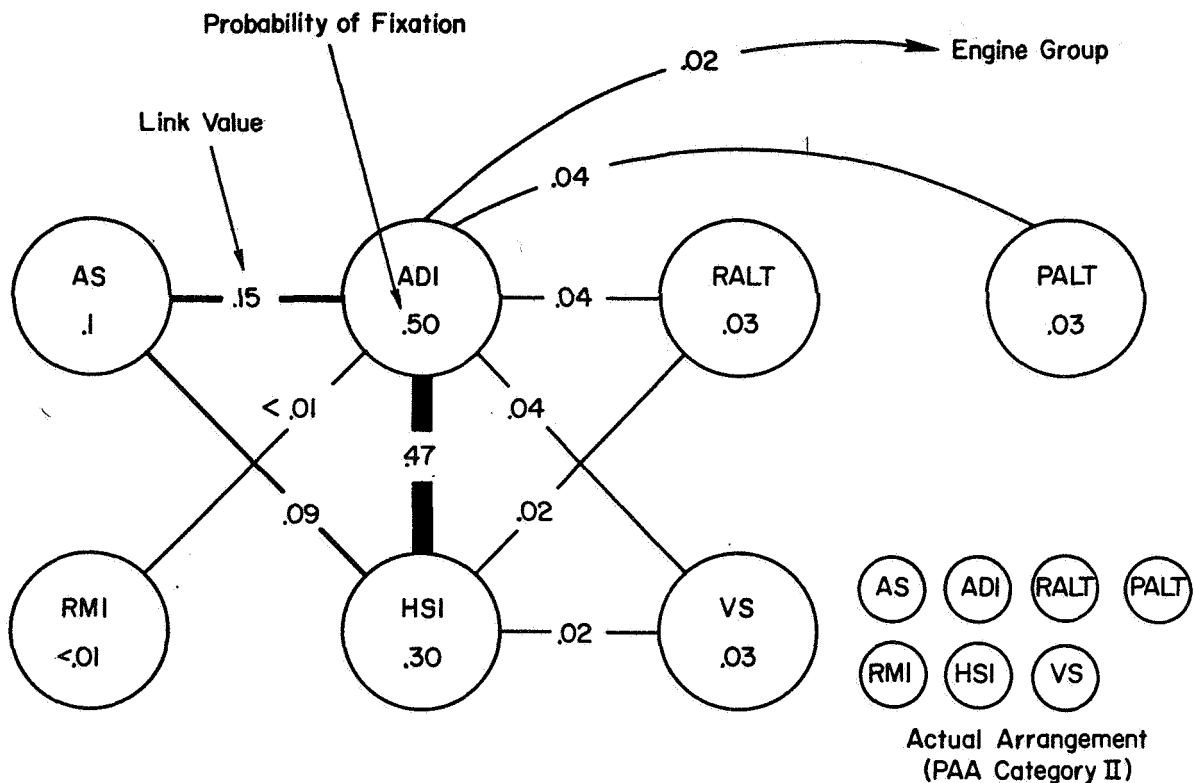


Figure 11. —Predicted display arrangement for B707-320 in ILS approach (for PAA category II).

Besides the obvious task of correlating these numerical predictions with actual piloted jet transport instrument landing data (which will be attempted later this year using the NASA Ames landing simulator). There are several elements of the theory needing refined models and a broader empirical data base:

- (1) Measurement of feedback selection, pilot describing functions and remnant under freely scanned multidisplay conditions (especially those with several instruments and loop closures potentially available)
- (2) Scanning and reconstruction models and penalties
- (3) More data on display workload (as defined herein), based on eye fixation data in a wide range of situations
- (4) A better measure of sensory loading (introduced in reference 2, but not satisfactorily formulated yet)
- (5) Analyses of other case studies (to validate and refine the theory)

Since several investigators are known to be working on these problem areas, we should see rapid progress during this next year. Once this theoretical-empirical framework is available, it will provide a sound and economic basis for: display development, experimental design and analysis, comparative evaluation of display systems, and generation of new display concepts.

Unfortunately, to perform a valid and comprehensive display analysis presently requires an awesome array of models (each with variable constants), wide working experience in several areas of systems engineering and psychophysiology, and an orgy of iterative analysis. A similarly discouraging situation existed several years ago in regard to multi-loop handling qualities, but it was overcome by more sophisticated techniques, interpretation of numerical results, and by use of more rapid computational schemes. The biggest challenge ahead for the systems analysis theory of manual control displays will be to simplify its application without losing its breadth or validity.

REFERENCES

1. McRuer, D. T.; and Jex, H. R.: A Systems Analysis Theory of Manual Control Displays. Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967, pp. 9-28.
2. McRuer, Duane; Jex, Henry R.; Clement, Warren F.; and Graham, Dunstan: Development of a Systems Analysis Theory of Manual Control Displays. Systems Technology, Inc., Tech. Rept. 163-1, Oct. 1967.
3. Weir, D. H.: Compilation and Analysis of Flight Control System Command Inputs. AFFDL-TR-65-119, Jan. 1966.
4. Houbolt, J. C.; Steiner, R.; and Pratt, J. G.: Dynamic Response of Airplanes to Atmospheric Turbulence, including Flight Data on Input and Response. NASA TR R-199, June 1964.
5. Lappe, U. O.; Thuillier, R. H.; and Reeves, R. W.: Development of a Low Altitude Turbulence Model for Estimating Gust Loads on Aircraft. ASD TDR-63-318, July 1963.
6. Pritchard, F. E.: A Statistical Model of Atmospheric Turbulence and a Review of the Assumptions Necessary for its Use. Cornell Aeronautical Laboratory, Inc. Presented at the AGARD Specialists' Meeting on Stability and Control (Cambridge, England), Sept. 20-23, 1966.
7. Eggleston, J. M.; and Diederich, F. W.: Theoretical Calculation of the Power Spectra of the Rolling and Yawing Moments on a Wing in Random Turbulence. NACA TN-3864, Dec. 1956.
8. Hart, J. E.; Adkins, L. A.; and Lacau, L. L.: Stochastic Disturbance Data for Flight Control System Analysis. ASD-TDR-62-347, Sept. 1962.
9. Teper, Gary L.; and Stapleford, Robert L.: An Assessment of the Lateral-Directional Handling Qualities of a Large Aircraft in the Landing Approach. J. Aircraft, vol. 3, no. 3, May-June 1966, pp. 201-207.
10. Clement, W. F.; Best, J. J.; and Stapleford, R. L.: Analysis of Display-Pilot-Vehicle Dynamics for ILS Approach of a Large Jet Aircraft. Systems Technology, Inc., TM 163-C, May 1967.
11. McRuer, Duane T.; and Jex, Henry R.: A Review of Quasi-Linear Pilot Models. IEEE Trans., vol. HFE-8, no. 3, Sept. 1967, pp. 231-249.
12. Bergen, A. R.: On the Statistical Design of Linear Random Sampling Schemes. Proc. IFAC. Vol. I, Butterworth, London, 1961, pp. 430-436.
13. Clement, Warren F.: Cardinal Reconstruction Theory: A Tool for Estimating Effects of Display Scanning. Systems Technology, Inc., TM-163-B, 1 Mar. 1967.
14. Stapleford, Robert L.; Craig, Samuel J.; and Tennant, J. A.: Measurement of Pilot Describing Functions in Single-Controller Multiloop Tasks. Systems Technology, Inc., Tech. Rept. 167-1, Aug. 1967.

15. Levison, W. H.; and Elkind, J. I.: Studies of Multivariable Manual Control Systems: Two-Axis Compensatory Systems with Separated Displays and Controls. NASA CR-875, Oct. 1967.
16. Stapleford, R. L.; and Ashkenas, I. L.: Effects of Manual Altitude Control and Other Factors on Short-Period Handling Quality Requirements. J. Aircraft, vol. 5, no. 1, Jan.-Feb. 1968, pp. 41-48.
17. Durand, Tulvio S.; and Wasicko, Richard J.: Factors Influencing Glide Path Control in Carrier Landing. J. Aircraft, vol. 4, no. 2, Mar.-Apr. 1967, pp. 146-158.
18. Clement, W. F.; Graham, D.; and Best, J. J.: A Re-examination of Eye Movement Data. Systems Technology, Inc., TM 163-A, 28 Feb. 1967.
19. Senders, J. W.; Elkind, J. I.; Grignetti, M.C.; and Smallwood, R.: An Investigation of the Visual Sampling Behavior of Human Observers. NASA CR-434, Apr. 1966.
20. Senders, J. W.: A Re-analysis of the Pilot Eye Movement Data. IEEE Trans., vol. HFE-7, no. 2, June 1966, pp. 103-106.
21. Jones, R. E.; Milton, J. L.; and Fitts, P. M.: Eye Fixations of Aircraft Pilots: I. A Review of Prior Eye Movement Studies and a Description of a Technique for Recording the Frequency, Duration, and Sequences of Eye Fixations During Instrument Flight, USAF AMC TR No. 5837, Sept. 1949.
22. Collins, Robert C.: The Different Views of Cat. II. AIAA Paper No. 65-766, Nov. 1965.
23. McRuer, D.T.; and Krendel, E. S.: The Man-Machine System Concept. Proc. IRE, vol. 50, May 1962, pp. 117-1123.
24. McGehee, W.: Comparative Study of Pilot Fatigue Resulting from Extended Instrument Flight Using the Standard AAF and British Instrument Panels. Final Rept., Project TED ATLR-601, US Navy, 1944.

7. An Analysis of Pilot Adaptation in a Simulated Multiloop VTOL Hovering Task

*E. W. Vinje and D. P. Miller
United Aircraft Corporation Research Laboratories*

Human pilot adaptation in a simulated multiloop VTOL (vertical takeoff and landing) hovering task was investigated using a series-loop closure model. Using the model equations, the pilot model adaptable parameters were computed from rms hovering performance data measured in flight simulator experiments for a variety of VTOL aircraft configurations. Variations in the aircraft stability parameters affected both the dynamic and disturbance-response characteristics of the controlled element.

The pilot's pitch-loop adaptation generally correlated with the frequency domain characteristics of the pitch response to turbulence and the pitch response to control inputs. Gain was adapted to provide a crossover frequency that was high enough to control the disturbances due to turbulence and provide satisfactory hovering performance. Lead compensation was adapted in the region of crossover frequency to provide closed-loop stability. The results of the study show no low-frequency attitude gain requirements that must be satisfied for the attitude disturbance and dynamic characteristics considered. The influence of the aircraft position-loop characteristics on pilot pitch-loop adaptation was important only for configurations that had small attitude disturbances. For these cases the requirements for position-loop control apparently determined the minimum acceptable pitch-loop crossover frequency. Results also show that the pilot preferred to adjust pitch-loop gain by changing control stick sensitivity (aircraft gain) rather than by changing his internal adapted gain.

INTRODUCTION

Through intelligent design and application of stability augmentation systems, advanced vehicles such as VTOL aircraft, lifting reentry bodies, orbiting space vehicles, and lunar landing modules can be made to have control characteristics that are acceptable to most pilots. However, the margin for error in initial stability augmentation system design decreases as the flight vehicles and tasks become more complex. A better understanding of human pilot control characteristics is necessary. In particular, a practical mathematical definition of human pilot control characteristics would be an asset to the designer.

A number of studies have been conducted with the objective of mathematically modeling pilot control characteristics (for example, refs. 1, 2 and 3). From a practical engineering point of view, the development of the quasi-linear model for human pilot dynamics has been the most beneficial outcome of these studies (ref. 1). The quasi-linear model describes pilot dynamic control characteristics in terms of frequency-domain servoanalysis parameters, i.e., adaptable gains, leads and lags, and a relatively constant neuromuscular lag and transport lag. Qualitative rules defining the proper adjustment of the adaptable model parameters to simulate pilot adaptation have also been developed in these studies.

In the development of the quasi-linear pilot models, most studies have been limited to pilot control in a one-degree-of-freedom task. In many flight vehicles, however, the most fundamental task is the control of two coupled degrees of freedom. An example of this is the control of pitch attitude and longitudinal hovering position for a VTOL aircraft. Because the pilot is limited in his control capabilities, one would not necessarily expect that the rules governing pilot adaptation in one-degree-of-freedom control tasks would also apply to two-degree-of-freedom control tasks. Pilot adaptation in two-degree-of-freedom tasks is, therefore, an area that requires additional study.

A previous study of the characteristics of pilot dynamics in a two-degree-of-freedom task was conducted at the United Aircraft Research Laboratories (UARL). The objective was to determine the correlation between pilot opinion, optimum control sensitivity, and computed pilot adapted parameters (ref. 4). A two-degree-of-freedom simulated VTOL aircraft hovering task was modeled as a multi-closed-loop system in which pilot dynamics were represented with quasi-linear models. The pilot adapted parameters (gain and lead) were computed from rms hovering performance data measured in the UARL V/STOL aircraft flight simulator. Pilot opinion was found to correlate with pilot adapted lead and pilot workload (as indicated by the aircraft response to turbulence). The pilot's tendency to maintain constant adapted pitch-loop gain for a wide range of configuration parameters apparently was the criterion used to select optimum control sensitivity. The results of the aforementioned study suggest the possibility of analytically predicting pilot opinion and optimum control sensitivity. If the adapted pilot model parameters and the characteristics of aircraft response to turbulence can be estimated, an estimate of pilot opinion and optimum control sensitivity might be made. However, the usefulness of this analytical approach depends upon the reliability of the estimated pilot-adapted parameters. The criterion governing pilot adaptation in multiloop control tasks must be understood to properly define these parameters.

This present study was undertaken to explore the factors that influence pilot adaptation for a simulated two-degree-of-freedom hovering task in turbulent air.

SYMBOLS

dB	decibels, $20 \log_{10} ()$ where () is an amplitude quantity, e.g., $\theta(s)/\delta_e(s)$
g	gravitational constant, 32.2 ft/sec ²
I_y	moment of inertia in pitch, slug-ft ²
j	$\sqrt{-1}$
$K_{P\theta}$	pilot adapted pitch-loop gain, in./rad
K_{Px}	pilot adapted position-loop gain, rad/ft
m	mass of aircraft, slugs

M_q	pitching moment due to pitch rate divided by I_y , $1/\text{sec}^2$
M_u	pitching moment due to longitudinal velocity divided by I_y , $1/(\text{ft-sec})$
M_{δ_e}	pitching moment due to unit longitudinal control stick displacement divided by I_y , $(\text{rad/sec}^2)/\text{in.}$
M_θ	pitching moment due to pitch attitude divided by I_y , $1/\text{sec}^2$
PM	phase margin, an indicator of closed-loop stability expressed as the phase lag at crossover frequency of the aircraft open-loop dynamics and the pilot model plus 180° , deg
q	pitch rate, rad/sec
s	Laplace operator, $1/\text{sec}$
S_{u_g}	turbulence power spectrum, ft^2/sec
T	constant defining first-order root location in the complex plane, sec
T_{L_x}	pilot adapted position-loop lead, sec
T_{L_θ}	pilot adapted pitch-loop lead, sec
T_N	neuromuscular lag, sec
u	velocity component in x direction of stability axis system, ft/sec
u_g	longitudinal component of gust velocity, ft/sec
x	longitudinal coordinate in stability-axis system, ft
X_u	longitudinal force due to longitudinal velocity divided by m , $1/\text{sec}$
Y_{P_θ}	pitch-loop quasi-linear pilot model (see eq. (3)), $\text{in.}/\text{rad}$
Y_{P_x}	position-loop quasi-linear pilot model, rad/ft
δ_e	longitudinal displacement of control stick from center of travel, in.
ζ	damping ratio of oscillatory roots (see eqs. (4) and (5))
θ	pitch attitude relative to horizon, rad
θ_x	position-loop command angle, rad (see fig. 2)
$\theta(s)/u_g(s)$	transfer function relating pitch attitude to turbulence (see table 1), $\text{rad-sec}/\text{ft}$
$\theta(s)/\delta_e(s)$	transfer function relating pitch attitude to control inputs (see table 1), $\text{rad}/\text{in.}$
σ_q	RMS pitch rate, rad/sec
σ_u	RMS longitudinal velocity, ft/sec
σ_{u_g}	RMS longitudinal turbulence, ft/sec
σ_x	RMS longitudinal hovering position error, ft
σ_θ	RMS pitch attitude, rad
τ_θ	pitch-loop pilot model transport lag, sec

τ_x	position-loop pilot model transport lag, sec
ω	frequency, rad/sec
ω_B	break frequency of first-order filter for generating turbulence, rad/sec
ω_d	damped frequency of oscillation, rad/sec
ω_n	natural frequency of oscillation, rad/sec

Superscript:

denotes first derivative with respect to time, d/dt

Subscripts:

REF denotes reference or base value of a variable

x, θ denotes quantities pertaining to position and pitch loops, respectively

e denotes error between desired and actual value of a variable

SIMULATED VTOL HOVERING TASK

Brief discussions of the equations of motion, the simulated turbulence, the UARL V/STOL aircraft flight simulator, and methods used in performing the experimental portion of this study are presented here. More detailed discussions are contained in reference 4.

EQUATIONS OF MOTION.—The general form of the linearized equations of motion that describe the pitch attitude (θ) and longitudinal position (x) responses of the simulated hovering VTOL aircraft to control inputs and to turbulence are

$$\begin{aligned} M_u \dot{u} + M_\theta \dot{\theta} + M_q \dot{\theta}^I &= -M_{\delta_e} \delta_e - M_u u_g \\ X_u u - g \cos \theta - \dot{u} &= -X_u u_g \end{aligned} \quad (1)$$

The left side of the equations defines the aircraft transient response (aircraft dynamics). The right side defines the manner in which control inputs (δ_e) and turbulence (u_g) force aircraft motion. The equations have been normalized with mass and moment of inertia. These equations were used to develop the transfer functions for the closed-loop model and were also programed on the analog computer for use with the flight simulator.

The stability derivatives M_u (generally presented as M_{u_g}), M_q , and X_u all affect the aircraft's dynamic response. Also, the small rapid changes in airspeed resulting from turbulence cause pitching motion through M_{u_g} and displacement from the hovering position through X_u . Pitch rate damping, M_q , reduces the aircraft's pitch response to turbulence by developing a restoring moment proportional to pitch rate. Pitch attitude stabilization, M_θ , causes a restoring moment proportional to pitch attitude. Longitudinal control sensitivity, M_{δ_e} , is the initial pitch acceleration per inch of control stick displacement.

SIMULATION OF TURBULENCE.—The simulated turbulence was generated by passing the output of a low-frequency random noise generator through a first-order filter with a break-point frequency, ω_B , of 0.314 rad/sec. The resulting simulated turbulence had a power spectral density given by

$$S_{u_g}(\omega) = \frac{2\omega_B \sigma_{u_g}^2}{\omega_B^2 + \omega^2} \quad (2)$$

where σ_{u_g} is the root mean square value of the random component of the wind. A value of $\sigma_{u_g} = 5.14$ ft/sec was used for this study. In a previous flight test study (ref. 5) the pilots described simulated turbulence having these characteristics as corresponding to a medium-turbulent day.

UARL V/STOL AIRCRAFT FLIGHT SIMULATOR.—This is a full-scale fixed-base Sikorsky S-61 helicopter cockpit with a Norden Contact Analog display system (ref. 4). Aircraft pitch attitude was controlled with a conventional helicopter control stick having inertia but no force gradient. Longitudinal position was controlled through longitudinal forces produced by changing pitch attitude to rotate the thrust vector. .

The Norden Contact Analog system provided the pilot with an inside-out video picture of the motion of the aircraft relative to the Earth and sky. Figure 1 shows the display (ref. 4). The cross indicated the position of the nose of the aircraft relative to the ground and hence remained fixed at the center of the viewing screen. The square traversed

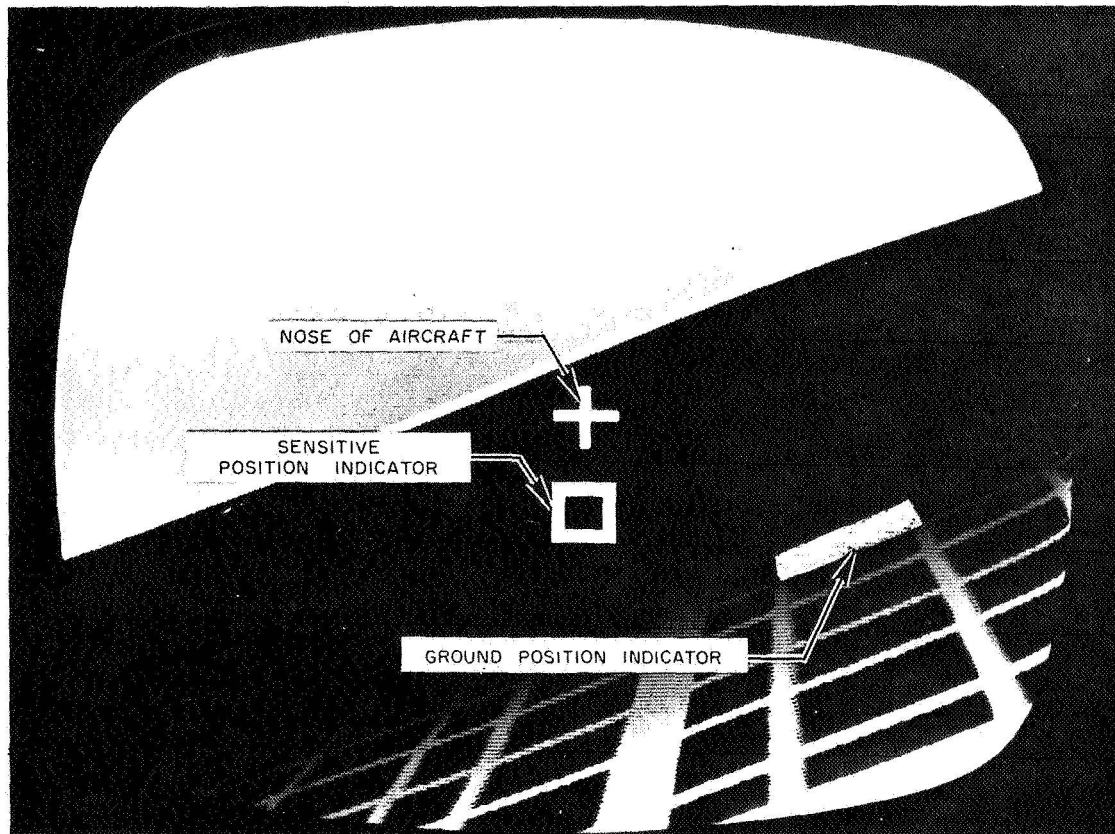


Figure 1.—Contact analog display for hovering and low-speed maneuvering task.

vertically on the screen and was used as a sensitive longitudinal position indicator. When the square was at the center of the screen (coincident with the cross) the aircraft was located directly over a reference hovering position such that the ground position indicator (also shown in fig. 1) was approximately 135 feet ahead of the aircraft for the simulated hovering altitude of 40 feet. Figure 1 shows the aircraft rolled to the right. The roll-attitude and lateral-position degrees of freedom were included in the simulated task. However, the roll motion was well damped and lateral disturbances caused by turbulence were small so that little effort was required to control roll attitude and lateral position (lateral control characteristics were not changed for any part of the study). As a result, the presence of the lateral degrees of freedom did not appreciably affect the pilot's performance of the longitudinal hovering task.

INSTRUCTIONS TO PILOT.—The pilot performed the hovering task for ranges of values of M_{ug} , M_q , and M_θ . The ranges investigated were:

$$0 \leq M_{ug} \leq 1.00$$

$$-5 \leq M_q \leq -1$$

$$-5 \leq M_\theta \leq 0$$

For each configuration the pilot was instructed to select the value of control sensitivity that he considered optimum for the hovering task. Control sensitivity was adjusted using a lever in the cockpit. He then attempted to minimize longitudinal hovering position error; i.e., he attempted to keep the cross centered in the square. The pilot performed this task for 10 periods of 100-second duration each. For each of these periods, rms longitudinal hovering position error, σ_x , rms longitudinal velocity, σ_u , rms pitch attitude, σ_θ , and rms pitch rate, σ_q , were measured. As will be discussed later, these rms values were subsequently used in computing the adapted pilot model parameters.

HOVERING TASK MODEL AND METHOD FOR COMPUTING PILOT ADAPTATION

This section contains a brief description of the model used in the analysis of pilot adaptation and the method for computing pilot adaptation. The mathematical representation for the human pilot, the assumed loop-closure structure and the computation of changes in pilot adaptable parameters from measured data are discussed. More detailed discussions of the implications of the task model and the technique used to compute pilot adaptation are contained in reference 4.

HUMAN PILOT MODEL.—The models selected to represent the human pilot dynamics (fig. 2) were linear except for a pure time delay or transport lag. These quasi-linear models have been found to be realistic approximations for pilot dynamic behavior in both one- and two-degree-of-freedom compensatory tracking tasks (refs. 1 and 2). The hovering task is also a compensatory tracking task. Thus, the basic form of these models can be used here as well.

The form of quasi-linear pilot model used for the pitch loop (fig. 2) is given in Laplace transform notation by

$$Y_{P_\theta} = \frac{K_{P_\theta} (T_{L_\theta} s + 1) e^{-\tau_\theta s}}{T_N s + 1} \quad (3)$$

Pilot gain, K_{P_θ} , represents the magnitude of the pilot's control response to a unit error in the magnitude of the controlled variable. The lead term, $(T_{L_\theta} s + 1)$, is an indication of the pilot's control response to the rate of change in error of the controlled element variable. The parameters K_{P_θ} and T_{L_θ} are adaptable. When the aircraft dynamics and/or turbulence characteristics change, the pilot can adapt his gain and lead to the new conditions. Through this adaptation the pilot can generally maintain stable control and achieve desirable system response characteristics. The model factors $1/(T_N s + 1)$ and $e^{-\tau_\theta s}$ define inherent pilot lags. The transport lag constant, τ_θ , represents an accumulation of delays involved in the transmission of information and decisionmaking. The other lag constant, T_N , is an approximation to the neuromuscular lag involved in the response of the arm to a command. The neuromuscular lag term is only present in the pitch-loop pilot model. This is because the series-loop model assumes that arm activity is restricted to the pilot's control of the pitch loop. These lags and limitations in the assumption that they are constant are discussed in more detail in references 1 and 4. The values used in this study for these constants, which were chosen from the results of reference 1, were $\tau_\theta = 0.09$ second and $T_N = 0.35$ second. The position-loop pilot model (fig. 2) contains terms similar to those described for the pitch-loop pilot model. However, as indicated previously, the only lag term present in Y_{P_x} is that resulting from the transport lag constant, τ_x .

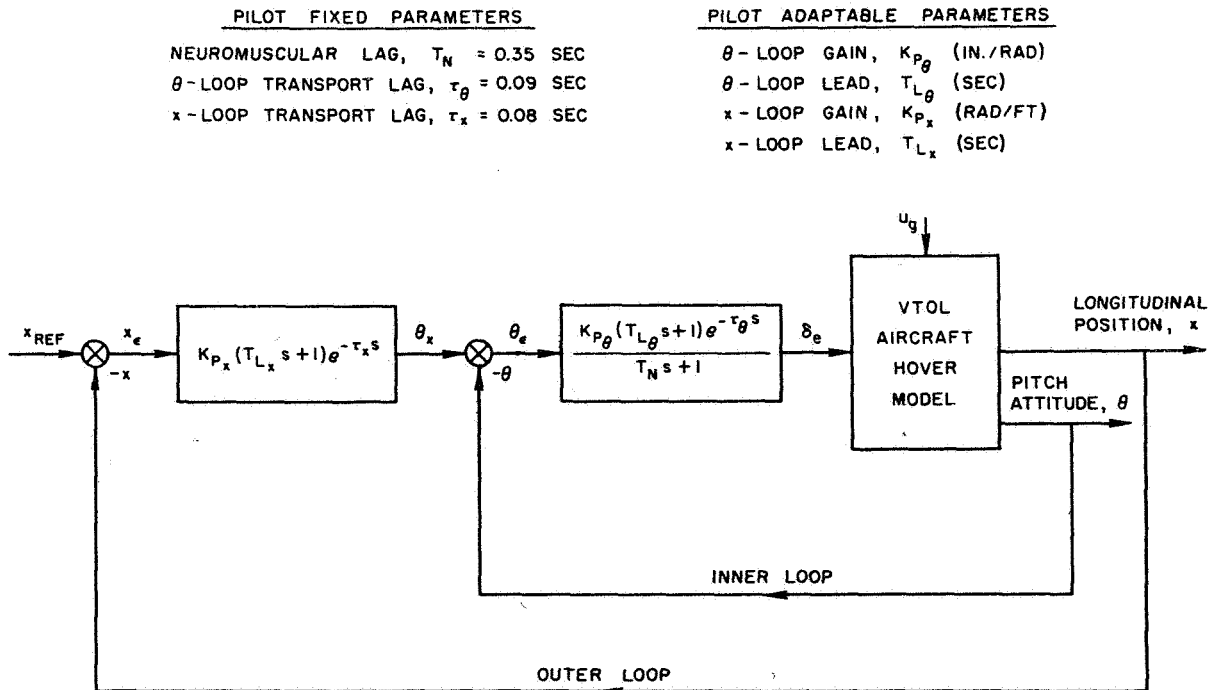


Figure 2.—Series-loop model for pilot control in hover.

LOOP-CLOSURE STRUCTURE.—The model in figure 2 assumes that the pilot closes the pitch (inner) loop and position (outer) loop in series. The pilot observes his hovering position error and mentally generates a pitch-attitude command angle, θ_x . The attitude command angle is that angle which will provide the component of horizontal thrust needed to develop the longitudinal acceleration required for reducing position error. This attitude angle command is routed to the pitch loop. There the pilot model develops a control input, δ_e , to null the error between the commanded attitude angle and the actual pitch attitude.

The series-loop closure is based upon observed pilot behavior. It was selected after first considering a parallel-loop closure model. The basic form of the series model was first brought to the attention of UARL by R. L. Stapleford of Systems Technology, Inc. An identical model has been developed at MIT for a helicopter longitudinal two-degree-of-freedom hovering task (ref. 6).

COMPUTATION OF PILOT-ADAPTED PARAMETERS.—Previous studies concerned with the development of quasi-linear pilot models (refs. 1 and 2) have involved measurement of random-excitation describing functions to represent pilot dynamic behavior. This permits evaluation of (1) the effect of the controlled element dynamics on the pilot's adaptation, (2) the extent to which he is a linear operator, and (3) the effects of his inherent nonlinearities on his adaptation. However, the procedures necessary to compute describing functions are complex and costly, especially for multi-degree-of-freedom control tasks.

Since the objectives of the UARL study were limited to determining trends in pilot-adapted parameters with changes in aircraft stability parameters, a different method was developed. It involves iterating on the pilot-model adapted parameters until the computed rms performance of the closed-loop hovering model matches measured rms hovering performance data obtained in the flight simulator studies. Details of the computational method, its verification, and the hovering task model equations are contained in reference 4. In addition, reference 4 contains a discussion of the effects of pilot nonlinearities on the computed parameters and how these effects are reduced by judicious selection of the rms performance data.

EFFECT OF AIRCRAFT STABILITY PARAMETERS ON PILOT ADAPTATION

Results are presented here to show the effects of changes in the aircraft speed-stability, pitch rate damping, and pitch attitude stabilization parameters on pilot adaptation. These parameters influence the aircraft dynamic response to control inputs and its pitch attitude response to the simulated turbulence. For each parameter, these response characteristics are presented in the frequency domain. The corresponding rms data from the flight simulator are also presented to show the effect of these parameters on hovering performance. The computed loop-closure characteristics (crossover frequencies and phase margins) and the pilot-adapted parameters needed to match the measured rms hovering performance data are discussed. In this paper, only the pitch-loop adaptation characteristics are discussed in detail. Also, results are presented for only one pilot. Data for a second pilot are presented in reference 4; the trends of the corresponding computed model parameters agree well with the results presented here.

LONGITUDINAL SPEED-STABILITY PARAMETER, M_{ug}

Open-Loop Response to Turbulence and Control Inputs.—Four values of longitudinal speed-stability parameter ($M_{ug} = 0, 0.33, 0.67, \text{ and } 1.00$) were considered. The other longitudinal stability derivatives remained unchanged. The simulated turbulence was filtered by the aircraft dynamics before it affected pitch attitude (fig. 2). This filtering effect can be described by $\theta(s)/u_g(s)$, the transfer function relating pitch attitude response to turbulence. The general form of this transfer function is given by

$$\frac{\theta(s)}{u_g(s)} = \frac{M_{ug}s}{\left(s + \frac{1}{T}\right)(s^2 + 2\zeta\omega_n s + \omega_n^2)} \quad (4)$$

The specific forms (in terms of the aircraft stability parameters) of all transfer functions discussed herein are given in table 1. As M_{ug} increases, the gain of $\theta(s)/u_g(s)$ at frequencies greater than 0.4 to 0.5 rad/sec increases (fig. 3a) and the pilot is required to attenuate disturbances having larger amplitudes. This, of course, increases the difficulty of the hovering control task. Note in equation (4) that for $M_{ug} = 0$, aircraft pitch attitude does not respond to turbulence.

Variations in M_{ug} also affect the aircraft's pitch response to control inputs, i.e., the dynamics of the controlled element change. The pitch response to control inputs is given by

$$\frac{\theta(s)}{\delta_e(s)} = \frac{M_{\delta_e}(s - x_u)}{\left(s + \frac{1}{T}\right)(s^2 + 2\zeta\omega_n s + \omega_n^2)} \quad (5)$$

For $M_{ug} = 0$, the denominator has three real roots (table 1). As M_{ug} increases, the roots change to a real root and a complex-conjugate pair. The natural frequency, ω_n , increases with M_{ug} and the damping ratio, ζ , decreases and becomes negative at $M_{ug} = 1.00$. The frequency response characteristics for control inputs $\theta(s)/\delta_e(s)$ are shown in figure 4 (M_{δ_e} has a constant value of 1.0 for these plots).

Measured RMS Hovering Performance.—An indication of the pilot's ability to adapt to changes in attitude disturbances and controlled element characteristics can be obtained by examining his rms hovering performance in the flight simulator experiments. These data are shown in figure 5a. Note that the pilot's rms hovering position error, hovering velocity, and pitch attitude changed very little with changes in M_{ug} . However, rms pitch rate increased with increasing M_{ug} . These results indicate that the pilot could attenuate relatively well the increased pitch attitude response to turbulence at the high values of M_{ug} . However, the pilot did not generate the increased pitch-loop lead (which will be shown later) needed to prevent an increase in rms pitch rate.

Loop-Closure Characteristics.—Specific closed-loop response and stability characteristics are required to match measured rms performance data with rms values computed from the equations for the series-loop model. The model closed-loop performance characteristics must correspond to those used by the human pilot in the performance of the task. A simplified means of describing the response and stability characteristics of the closed-loop

TABLE 1.—VTOL AIRCRAFT OPEN-LOOP LONGITUDINAL TRANSFER FUNCTIONS

Specific Form of the Transfer Functions	Configurations				Factors of Δ_1	
	M_q	X_u	$M_u g$	M_θ	$s + 1/T, s + \zeta\omega_n \pm j\omega_d$	
Δ_1	-3	-0.1	0.0	0	$s + 3.0, s, s + 0.10$	
			0.33		$s + 3.0, s + (0.03 \pm j 0.33)$	
			0.67		$s + 3.1, s + (0.01 \pm j 0.47)$	
			1.00		$s + 3.1, s - (0.004 \pm j 0.57)$	
e/δ_e						
e/u_g	-1	-0.1	0.67	0	$s + 1.4, s - (0.14 \pm j 0.68)$	
	-3				$s + 3.1, s + (0.013 \pm j 0.47)$	
x/δ_e	-5				$s + 5.0, s + (0.04 \pm j 0.36)$	
x/u_g	-1	-0.05	1.00	0	$s + 1.5, s - (0.21 \pm j 0.80)$	
				-3	$s + 0.4, s + (0.32 \pm j 1.64)$	
				-5	$s + 0.3, s + (0.40 \pm j 2.18)$	

$$\Delta_1 = s^3 - (M_q + X_u)s^2 + (M_q X_u - M_\theta)s + (M_u g + M_\theta X_u)$$

$$e/\delta_e = \frac{M_{\delta_e} (s - X_u)}{\Delta_1}$$

$$e/u_g = \frac{M_u s}{\Delta_1}$$

$$x/\delta_e = \frac{-M_{\delta_e} g}{s \Delta_1}$$

$$x/u_g = \frac{X_u (s^2 - M_q s - M_\theta) - M_u g}{s \Delta_1}$$

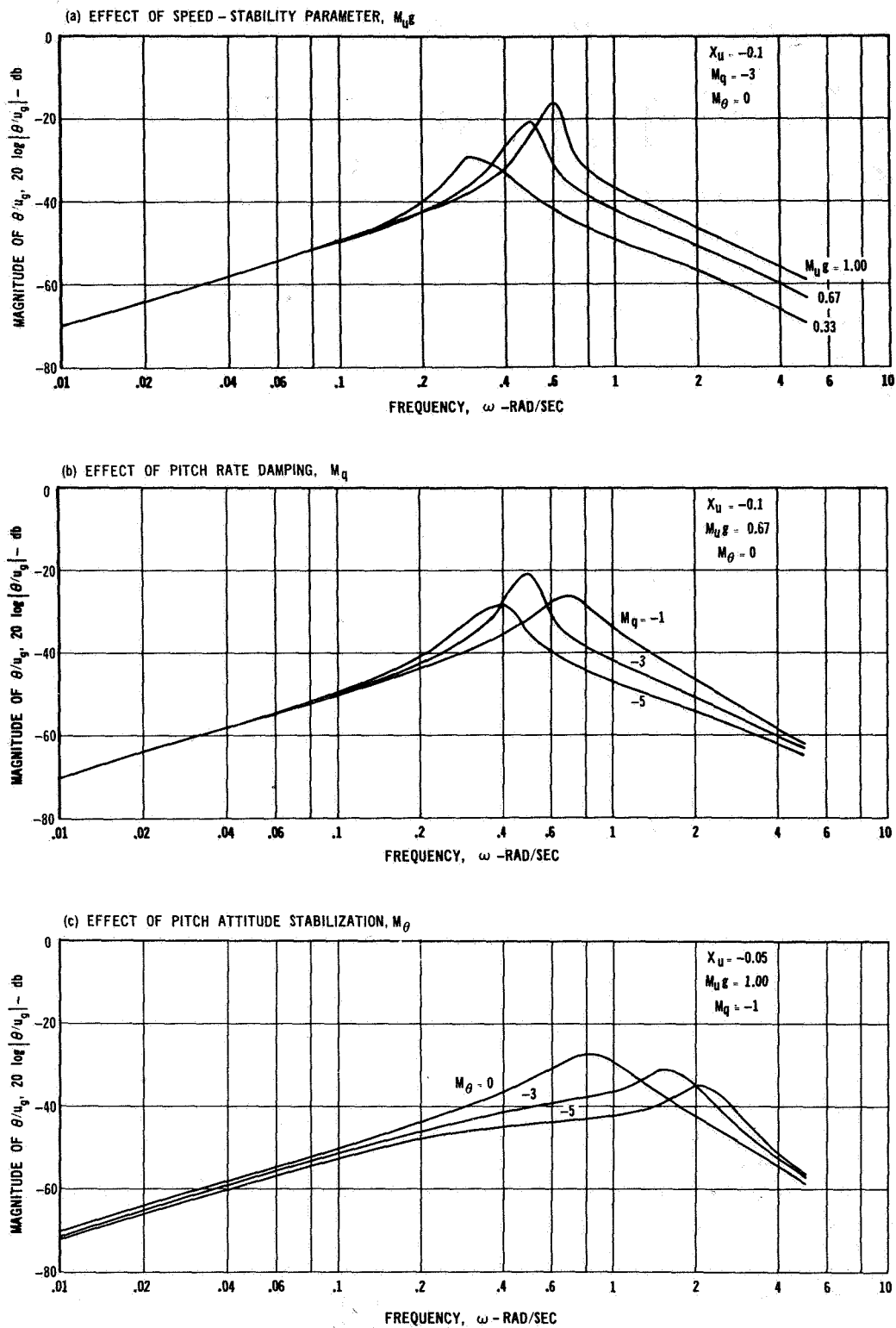


Figure 3.—Effect of aircraft stability parameters on the magnitude of the transfer function relating pitch attitude response to turbulence, $\theta(s)/u_g(s)$.

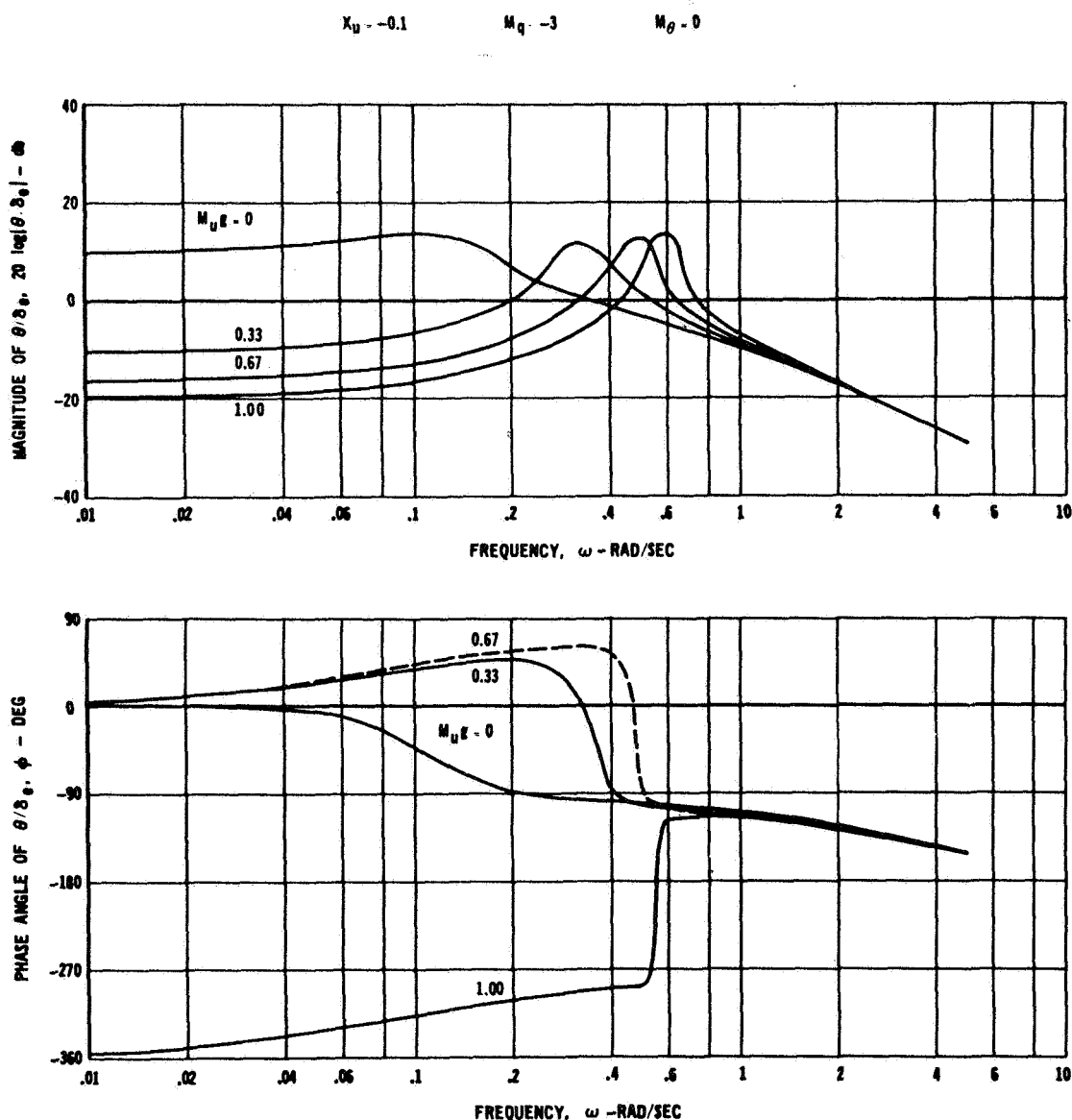


Figure 4.—Effect of speed stability parameter on the frequency response of the transfer function relating pitch attitude to control inputs, $\theta(s)/\delta_e(s)$.

model utilizes the crossover frequencies¹ (ω_c) and phase margins² (PM) of the two loops. The pitch-loop crossover frequency, $\omega_{c\theta}$, and phase margin, PM_θ , are obtained from Bode plots of the product of the transfer functions of the control elements included in the pitch attitude loop, $Y_{P\theta}(s)$ and $\theta(s)/\delta_e(s)$.

¹Pitch-loop crossover frequency, $\omega_{c\theta}$, is that frequency at which the gain of the product $Y_{P\theta}(s) \cdot \theta(s)/\delta_e(s)$ is equal to 1; increasing crossover frequency generally implies faster response to errors.

²Pitch-loop phase margin, PM_θ , is the sum of the phase angle of the aircraft open-loop dynamics and the phase angle of the pilot model at the crossover frequency, plus 180° ; the larger the phase margin the more stable the closed-loop response.

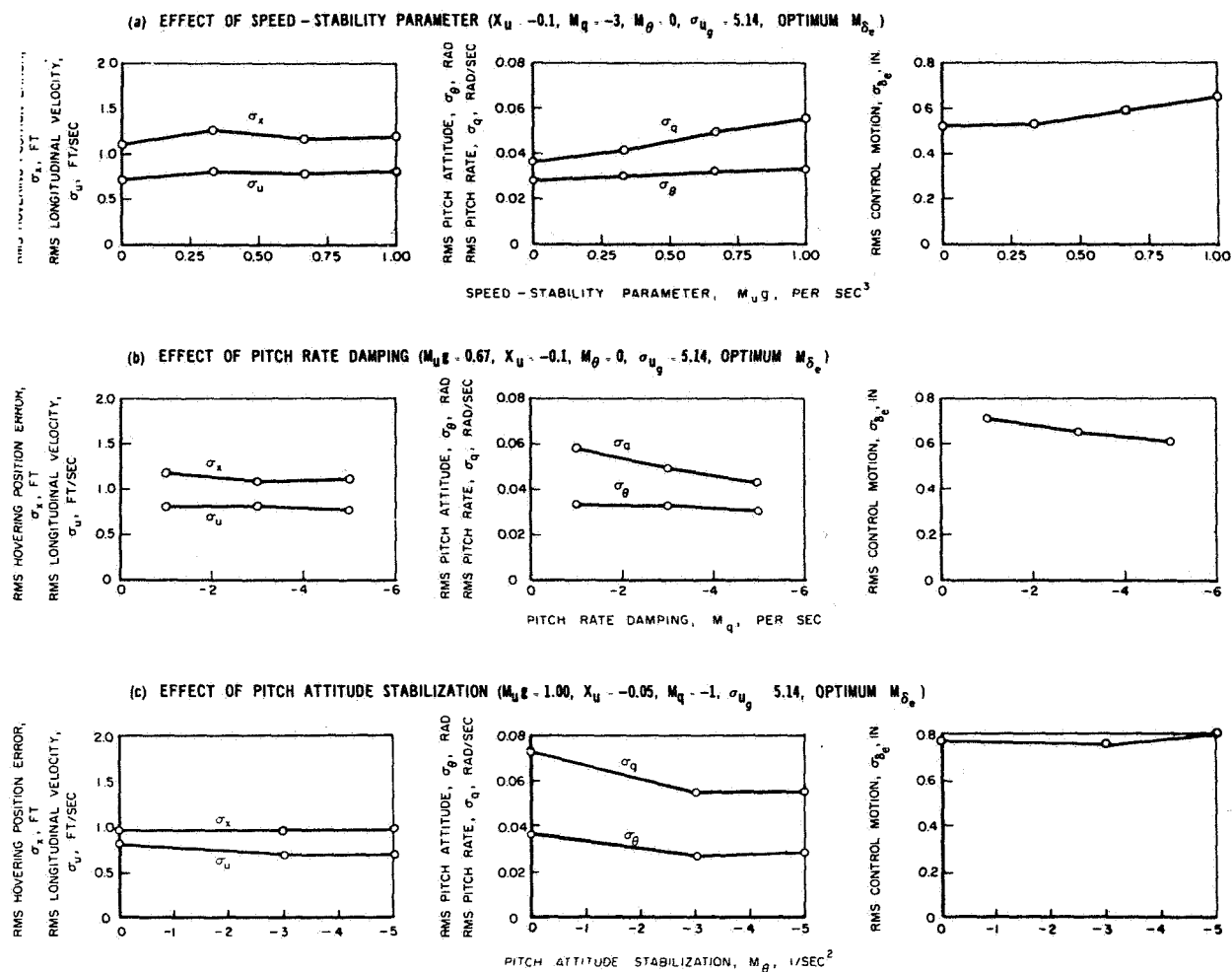


Figure 5. — Effect of aircraft stability parameters on RMS hovering performance.

Bode plots of $Y_{P_\theta}(s) \cdot \theta(s) / \delta_e(s)$ derived from the measured rms performance data for four values of M_{u_g} are shown in figure 6. Crossover frequency increased with increasing M_{u_g} . This is a reflection of the pilot's desire for faster pitch-loop response when attitude disturbances increased in amplitude at the higher frequencies (fig. 3a). The crossover frequencies and phase margins for the pitch and position loops were:

M_{u_g}	ω_{c_θ}	PM_θ	ω_{c_x}	PM_x
0	2.2	7	0.9	19
0.33	2.7	7	0.9	18
0.67	3.0	9	1.0	19
1.00	3.2	10	1.0	21

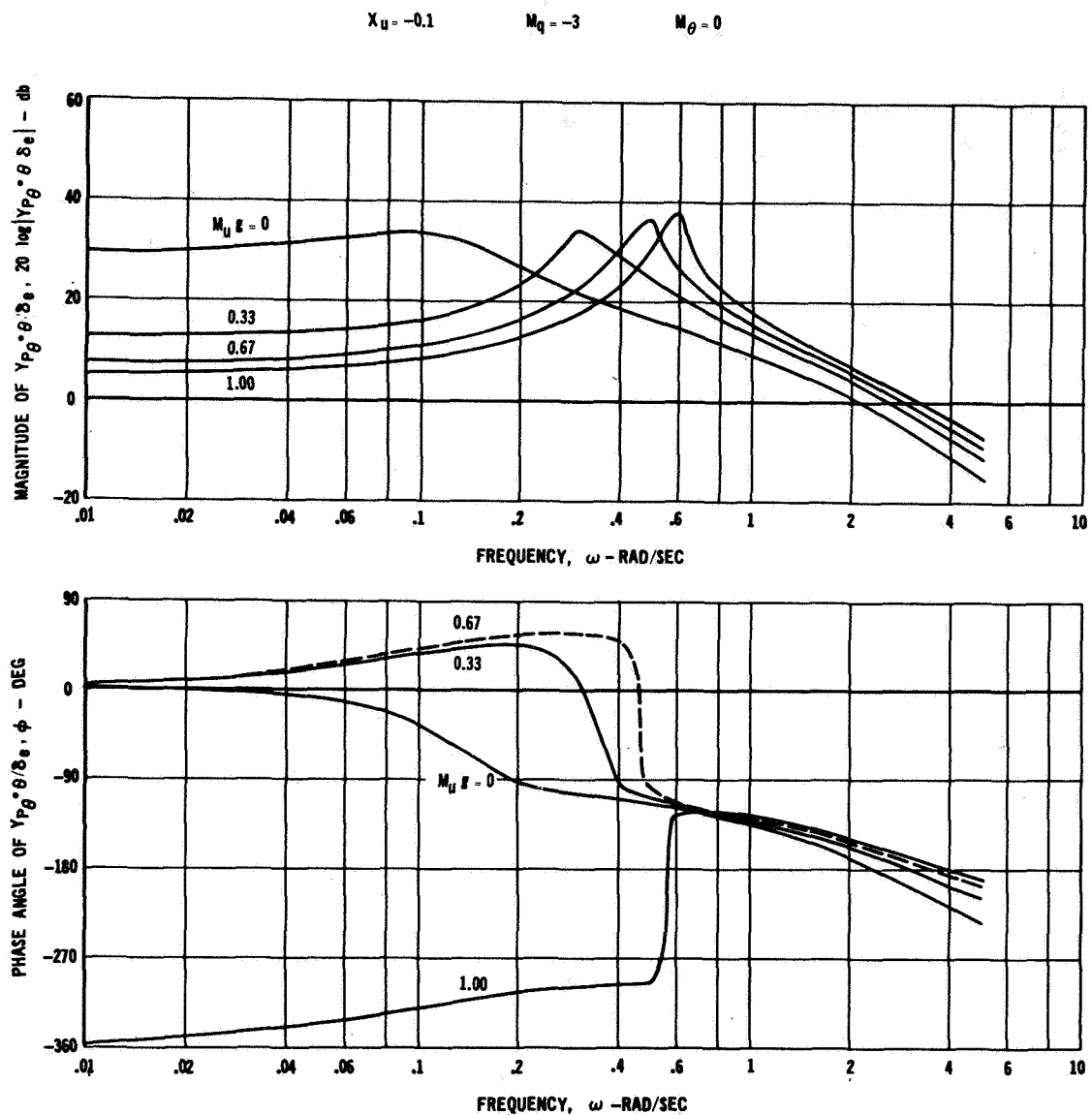


Figure 6.—Effect of speed stability parameter on the frequency response of the transfer function relating pitch attitude to pitch error, $Y_{P_\theta}(s) \cdot \theta(s) / \delta_e(s)$.

Note in figure 4 that the resonant frequency of $\theta(s) / \delta_e(s)$ increased with increasing M_{Ug} while the low-frequency gain ($\omega < 0.2$ rad/sec) decreased. Larger values of lead and gain had to be adapted in the pitch-loop pilot model at the higher values of M_{Ug} to supplement the response of $\theta(s) / \delta_e(s)$ and to provide the loop-closure characteristics. In terms of the actual pilot performance of the task, these computed loop-closure results imply that as M_{Ug} increased the pilot responded with larger control inputs for a given magnitude of observed error in pitch attitude. They also indicate that he correspondingly increased his anticipation of the effects of attitude rates to retain stability (increased ω_{c_θ} with nearly constant PM_θ). It should also be pointed out that for $M_{Ug} = 0$ (i.e., no pitch attitude disturbances due to

turbulence), $\omega_{c\theta}$ did not approach zero. Here, the pilot did not need the pitch control response to attenuate pitch disturbances, but he still had to command pitch attitude changes to reduce hovering position errors induced by position disturbances.

The closure of the position loop does not appear to have changed significantly with changes in M_{ug} . The small increase in ω_{cx} probably reflects the need for slightly faster position-loop response to attenuate the position-loop disturbances resulting from the small increase in σ_θ at the higher values of M_{ug} (fig. 5a).

Pilot-Adapted Parameters. —The pilot-adapted parameters required to provide the previously described loop-closure characteristics are shown in the following tabulation. The control stick sensitivity is also tabulated because it could be selected by the pilot, and thus can be considered an adaptable parameter.

$\underline{M_{ug}}$	$\underline{M_{\delta_e}}$	$\underline{K_{p\theta}}$	$\underline{M_{\delta_e} \cdot K_{p\theta}}$	$\underline{T_{L\theta}}$	$\underline{K_{px}}$	$\underline{T_{Lx}}$
0	0.300	32.6	9.8	0.006	-0.017	0.72
0.33	0.360	38.6	13.9	0.10	-0.020	0.53
0.67	0.431	37.7	16.2	0.16	-0.024	0.49
1.00	0.481	36.7	17.6	0.20	-0.0251	0.50

Adapted pitch-loop gain, $K_{p\theta}$, remained nearly constant with changes in M_{ug} while the pilot increased M_{δ_e} . Thus, the product $M_{\delta_e} \cdot K_{p\theta}$ increased with M_{ug} , indicating the need to increase the gain of $\theta(s)/\delta_e(s)$ and provide the pitch-loop crossover frequency required to suppress the high-frequency gust disturbances.

Low-frequency gain was required in the pitch loop to provide response to low-frequency pitch-attitude commands generated from observed errors in hovering position (θ_x in fig. 2). However, a large low-frequency gain was not an overriding requirement; in figure 6 it is shown that the pilot allowed the low-frequency gain of $Y_{p\theta} \cdot \theta(s)/\delta_e(s)$ to decrease with increasing M_{ug} while he increased $\omega_{c\theta}$.

These results also indicate that the pilot's criterion for selecting control sensitivity was to keep his internal gain constant while increasing the pitch-loop crossover frequency. That is, adapted $K_{p\theta}$ remained nearly constant while the pilot increased M_{δ_e} .

Adapted pitch-loop lead was small for all values of M_{ug} because the controlled element was relatively stable. The transient response from the first-order root in the controlled element dynamics dissipated quickly while the lightly damped oscillatory transient response was of such low frequency that the pilot had no difficulty in controlling it (table 1). The small increase in $T_{L\theta}$ with M_{ug} was necessary to extend $\omega_{c\theta}$ and maintain closed-loop stability at the higher values of M_{ug} .

The pilot's adapted position-loop parameters also changed with M_{ug} . However, these are more a reflection of the changes in the adapted pitch-loop parameters than an indication of any significant change in response to position-loop disturbances. This is because in the series-loop model the pitch-loop adapted parameters (M_{δ_e} , $K_{p\theta}$, $T_{L\theta}$) also affect the computed pilot response to position-loop disturbances (ref. 4). Consequently, if the pilot increases his response to pitch attitude disturbances (increasing $M_{\delta_e} \cdot K_{p\theta}$ and/or $T_{L\theta}$), and does not change his response to position disturbances, the computed position-loop adaptable parameters will change.

PITCH RATE DAMPING, M_q

Open-Loop Response to Turbulence and Control Inputs.—The effect of M_q on the Bode plots of the transfer function relating open-loop attitude response to turbulence is shown in figure 3b. As M_q becomes more negative, the resonant frequency decreases and the amplitude of the high-frequency ($\omega > 0.4$ to 0.5) attitude disturbances also decreases. For more negative M_q , then, aircraft pitch attitude becomes less responsive to high-frequency components of turbulence. The low-frequency attitude response to turbulence ($\omega < 0.2$) was essentially unchanged with changes in M_q . The effect of M_q on the roots is given in table 1.

The effect of M_q on the transfer function relating aircraft response to control inputs is shown in figure 7. The roots in table 1 show that the damped frequency of the aircraft dynamics decreases with more negative M_q while the damping ratio increases. The aircraft open-loop dynamics are unstable at $M_q = 1$, but stable at $M_q = -3$ and -5 . As shown in figure 7, at low frequencies ($\omega < 0.1$) the gain of $\theta(s)/\delta_e(s)$ is nearly invariant with M_q (again, M_{δ_e} has a value of 1.0 in these plots).

Measured RMS Hovering Performance.—The effect of M_q on rms hovering performance is shown in figure 5b. As was the case for changes in M_{ug} , σ_θ , σ_x , and σ_u were essentially unaffected by changes in M_q . The pilot was able to adapt to the lightly damped dynamics of the controlled element at $M_q = -1$, where the attitude response to turbulence at high frequencies was greatest, and to perform the hovering task as well as for the more heavily damped and stable dynamics at $M_q = -5$. Here, as in the case of changes in M_{ug} , the pilot was able to adapt to the variations in disturbance level and controlled element dynamics and maintain relatively constant hovering performance. Again, σ_q was largest for the most gust-sensitive configuration ($M_q = -1$). The pilot did not generate the pitch-loop lead ($T_{L\theta}$) necessary to maintain constant σ_q because he could maintain constant hovering performance without doing so.

Loop-Closure Characteristics.—The frequency response characteristics of the transfer function relating open-loop pitch attitude response to the observed pitch error, $Y_{P\theta} \cdot \theta(s)/\delta_e(s)$, are shown in figure 8 for the three values of M_q . The crossover frequencies and phase margins of this transfer function are shown in the following tabulation:

M_q	ω_{c_θ}	PM_θ	ω_{c_x}	PM_x
-1	3.4	8	1.0	21
-3	3.1	8	1.0	15
-5	2.8	10	1.0	19

As M_q became more negative and the pitch dynamics became more stable and less responsive to turbulence, pitch-loop crossover frequency decreased while the position-loop closure characteristics were essentially unchanged. As in the case of M_{ug} , the model results indicate that the pilot's main concern was to adjust his adaptable parameters to provide adequate response to the high-frequency disturbances ($\omega > 0.4$). He appeared less concerned about the low-frequency ($\omega < 0.3$) open-loop pitch gain because he tolerated significant differences in low-frequency gain for the different values of M_q . These results again would appear to reflect the relative unimportance of the low-frequency attitude disturbances.

However, some low-frequency gain and also some minimum value of pitch-loop crossover frequency must be maintained to provide acceptable pitch attitude response to correct for errors in hovering position.

The position-loop crossover frequency did not change with M_q . The pilot compensated for changes in pitch-loop characteristics in such a way that rms pitch attitude error remained nearly constant. Therefore, the contribution to the rms position error from the rms pitch attitude error was also nearly constant. The pilot therefore did not have to make any significant adjustments in his position-loop closure.

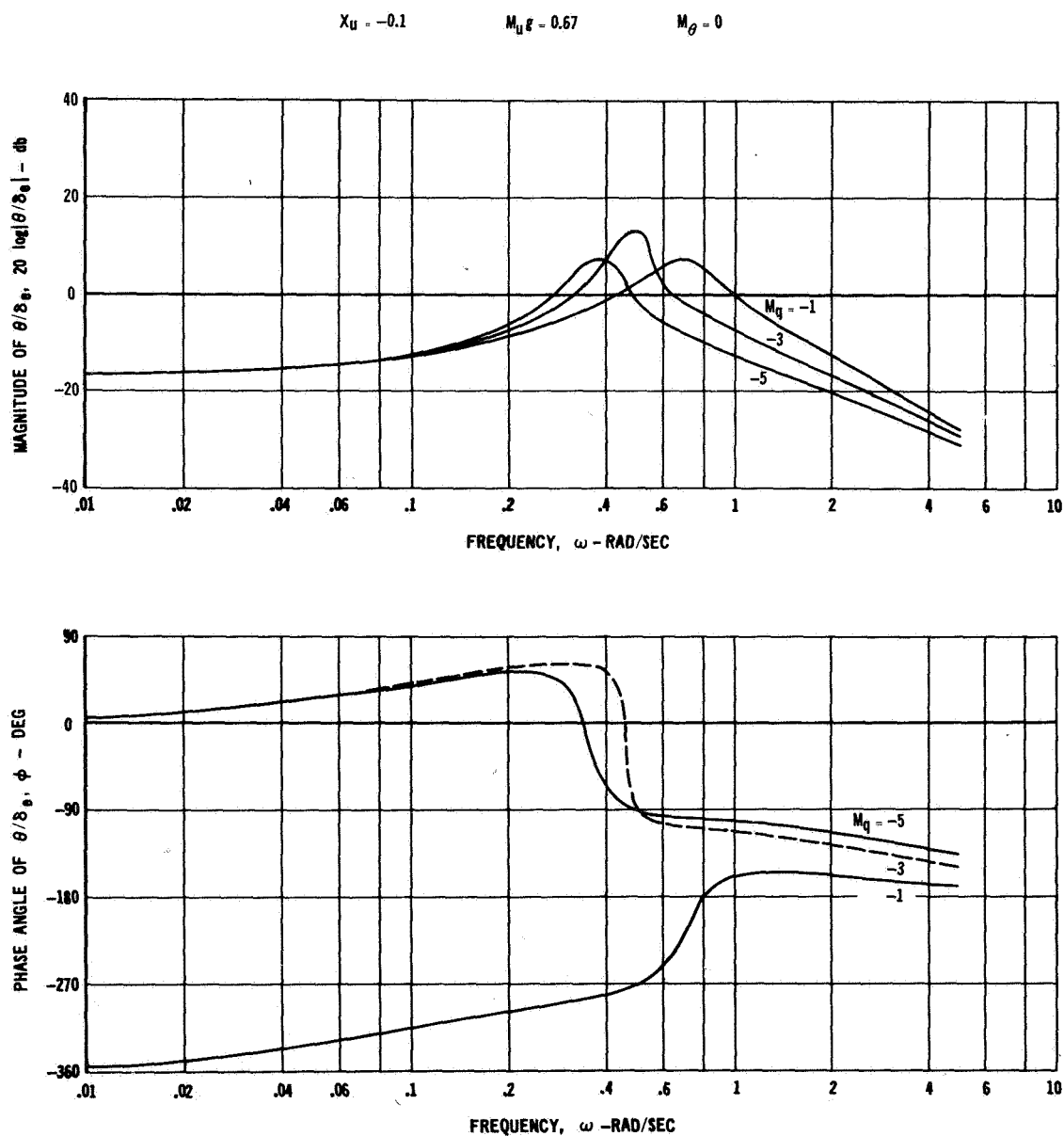


Figure 7.—Effect of pitch rate damping on the frequency response of the transfer function relating pitch attitude to control inputs, $\theta(s)/\delta_e(s)$.

Pilot-Adapted Parameters.—The pilot-model adaptable parameters are shown in the following tabulation:

$\underline{M_q}$	$\underline{M_{\delta e}}$	$\underline{K_{P_\theta}}$	$\underline{K_{P_\theta} \cdot M_{\delta e}}$	$\underline{T_{L_\theta}}$	$\underline{K_{P_x}}$	$\underline{T_{L_x}}$
-1	0.369	25.6	9.45	0.50	0.028	0.41
-3	0.431	39.7	17.06	0.17	0.027	0.41
-5	0.493	43.8	21.59	0.05	0.023	0.56

The product of pilot adaptable gain and control stick sensitivity, $K_{P_\theta} \cdot M_{\delta e}$, increased as M_q became more negative. This was required to compensate for the reduction in gain of $\theta(s)/\delta_e(s)$ at frequencies greater than about 0.4 rad/sec. The pilot adapted this increased gain even though the pitch attitude response to turbulence at these frequencies decreased. It should again be pointed out, however, that ω_{c_θ} became smaller as M_q became more negative. Also, the adapted gain resulted in larger low-frequency gain for $Y_{P_\theta} \cdot \theta(s)/\delta_e(s)$ as M_q became more negative (fig. 8). As discussed previously, the effects of M_q on the low-frequency characteristics of $\theta(s)/u_g(s)$ (the gain was small and remained unchanged with changes in M_q) and $\theta(s)/\delta_e(s)$ (gain also remained unchanged with changes in M_q) do not require the pilot to adjust his low-frequency gain. The effect of M_q on the low-frequency gain of $Y_{P_\theta} \cdot \theta(s)/\delta_e(s)$ would seem to be simply a result of the response requirements at higher frequencies and the form of the pilot model.

Adapted pitch-loop lead, T_{L_θ} , decreased as M_q became more negative. The controlled element dynamics became more stable as M_q became more negative. That is, the time constant of the first-order root decreased and the damping ratio of the oscillatory roots increased (table 1). The requirements for large ω_{c_θ} also decreased; consequently, there was a diminished requirement for T_{L_θ} .

The adapted position-loop parameters also changed with changes in M_q . However, as in the case of changes in M_{u_g} , these changes reflect the effect of the inner-loop changes rather than a fundamental change in the nature of the pilot's outer-loop control characteristics.

PITCH ATTITUDE STABILIZATION, M_θ

Open-Loop Response to Turbulence and Control Inputs.—The effect of M_θ on the pitch attitude response to turbulence is shown in figure 3c. As M_θ becomes more negative, there is a significant reduction in the frequency response to turbulence up to $\omega = 1$ rad/sec, but an increased response to turbulence at higher frequencies. The resonant frequency of the attitude response to turbulence increases for more negative M_θ but it appears that the over-all level of attitude disturbances (the total power) evident to the pilot is reduced.

The effect of M_θ on $\theta(s)/\delta_e(s)$ is shown in figure 9. The low-frequency gain decreases as M_θ becomes more negative and the high-frequency gain increases. The damped frequency of oscillation also increases as M_θ becomes more negative (table 1). The damping ratio increases when M_θ is changed from 0 (negative damping ratio) to -3, but decreases (remaining positive) again when M_θ is changed to -5. The aircraft dynamics were unstable at $M_\theta = 0$ and stable for $M_\theta = -3$ and -5.

Aircraft pitch attitude is difficult to control for $M_\theta = 0$ because it is unstable and quite responsive to turbulence. At $M_\theta = 0$ the time constant from the real pole is relatively

large and its effects are quite noticeable. Here, however, the damped frequency of the oscillatory pair is small and the transient effects resulting from the oscillatory dynamics are not particularly annoying. As M_θ is made more negative, the time constant of the real pole increases but its effect on aircraft transient response is canceled by the numerator term $(s - X_u)$ so that for $M_\theta = -3$ the dynamics are more easily controlled. However, as mentioned, the damped frequency of oscillation increases when M_θ becomes more negative. For $M_\theta = -5$ the higher oscillatory frequency and reduced damping ratio make control of the oscillatory portion of the dynamics somewhat difficult. The effect of M_θ on the root locations of the transfer functions $\theta(s)/u_g(s)$ and $\theta(s)/\delta_e(s)$ are shown in table 1 ($M_{\delta_e} = 1.0$).

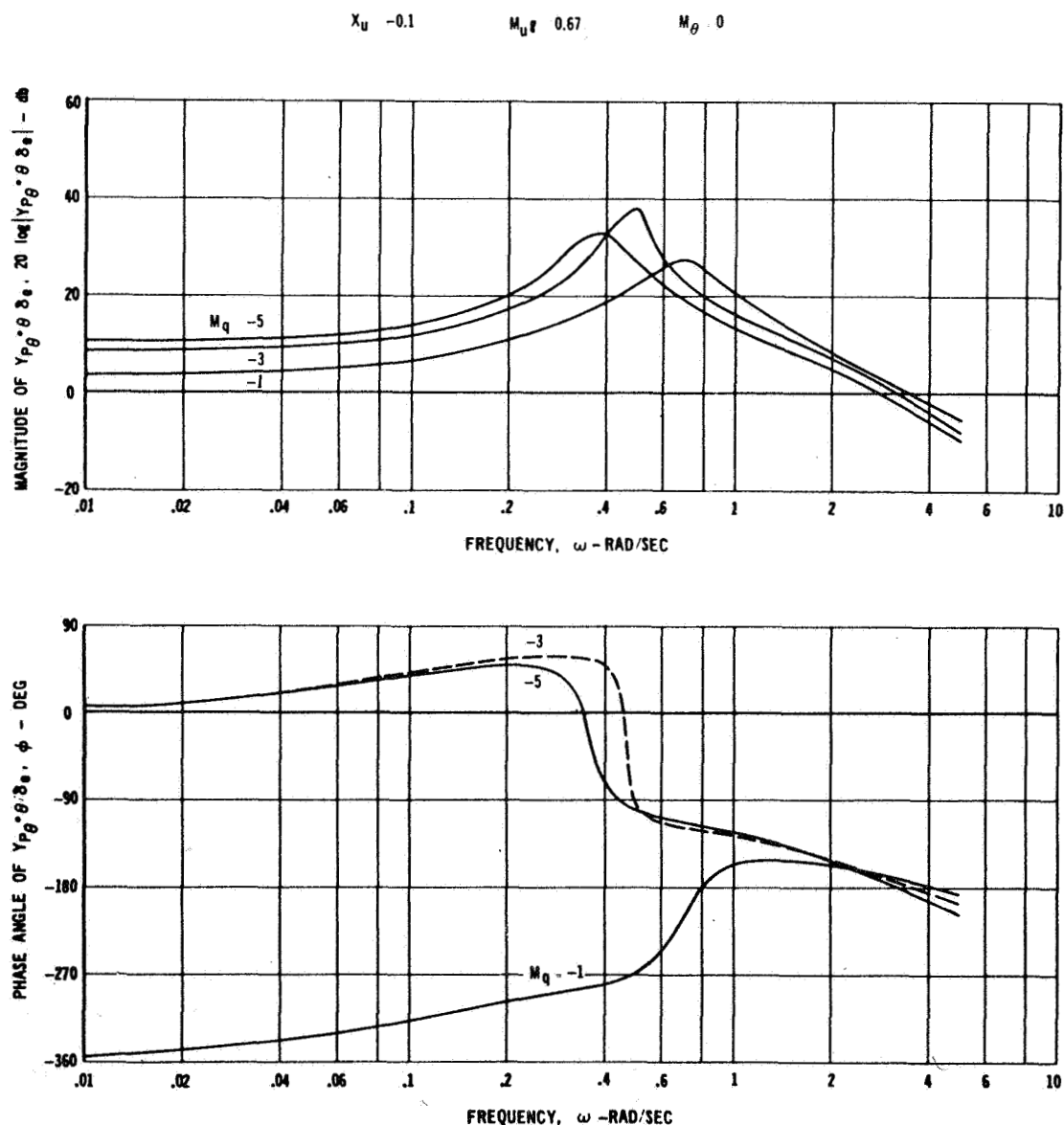


Figure 8. —Effect of pitch rate damping on the frequency response of the transfer function relating pitch attitude to pitch error, $Y_{P_\theta}(s) \cdot \theta(s)/\delta_e(s)$.

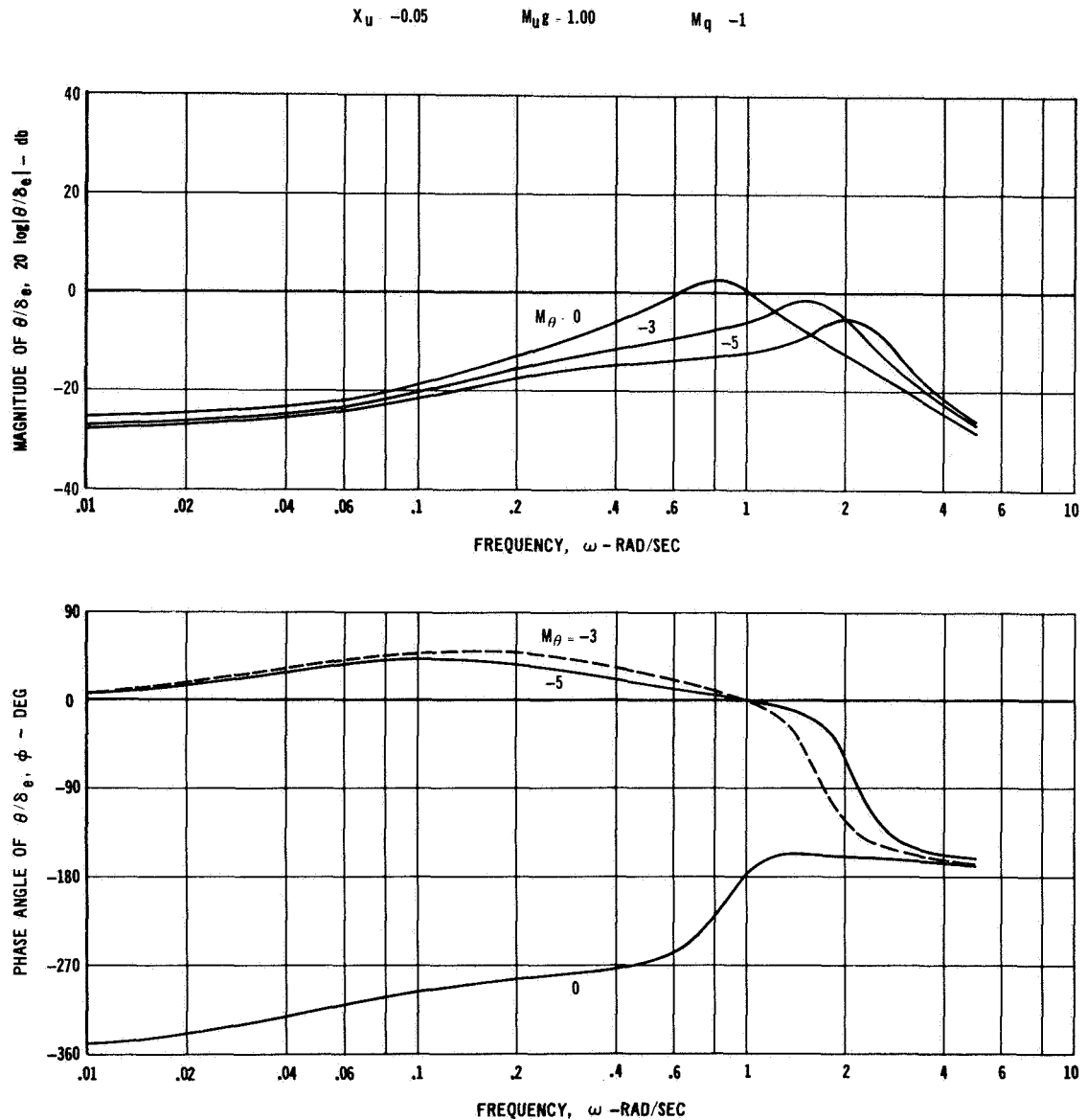


Figure 9.—Effect of pitch attitude stabilization on the frequency response of the transfer function relating pitch attitude to control inputs, $\theta(s)/\delta_e(s)$.

Measured RMS Hovering Performance.—The effect of M_θ on rms hovering performance is shown in figure 5c. As in the cases of changes in M_{ug} and M_q , the rms hovering position error and longitudinal velocity were essentially unaffected by M_θ . There was a reduction in both σ_θ and σ_q when M_θ was changed from 0 to -3. This reduction in σ_θ indicates that the pilot had some difficulty in suppressing attitude disturbances for $M_\theta=0$. When M_θ was made more negative, i.e., $M_\theta=-5$, there was little change in any of the rms values from those measured for $M_\theta=-3$. The effects of M_{ug} , M_q , and M_θ on the measured rms performance were quite similar. Measured values of σ_x and σ_u , and generally of σ_θ also, were relatively unchanged. However, σ_q (and σ_{δ_e}) increased for configurations that were more difficult to control and/or more responsive to turbulence.

Loop-Closure Characteristics.—A Bode plot for the transfer function $Y_{P_\theta}(s) \cdot \theta(s)/u_g(s)$ is shown in figure 10. The low-frequency gain of this transfer function ($\omega < 1$ rad/sec) decreased with more negative values of M_θ , while the high-frequency gain increased somewhat ($\omega > 1.5$ rad/sec). The crossover frequencies and phase margins computed from these plots are:

M_θ	ω_{c_θ}	PM_θ	ω_{c_x}	PM_x
0	3.2	11	1.2	17
-3	3.8	10	1.0	18
-5	3.6	13	1.0	18

Pitch-loop crossover frequency increased as M_θ was changed from 0 to -3 and then decreased slightly for $M_\theta = -5$. Even though the magnitude of the attitude disturbances over most of the frequency range was decreased by the addition of M_θ (fig. 3c), the computed pitch-loop crossover frequency increased. This apparently was an indication of the pilot's desire to attenuate the increased amplitude of the high-frequency disturbances for $M_\theta = -3$ and -5. However, ω_{c_θ} did decrease slightly at $M_\theta = -5$ when the total power of the attitude disturbances was decreased still further. The results shown in figure 10 again point out the apparent unimportance of the low-frequency pitch-loop gain to the pilot. The magnitude of the low-frequency pitch-loop gain to the pilot. The magnitude of the low-frequency attitude disturbances ($\omega < 0.1$) for M_θ differ little from those for M_{u_g} and M_q (fig. 3). However, the low-frequency gain of $Y_{P_\theta} \cdot \theta(s)/u_g(s)$ computed from the M_θ data is significantly lower than those computed using the M_{u_g} and M_q data for the same frequency region. This is consistent with the low amplitude of attitude disturbances over this frequency band. However, it might be supposed that some minimum value of low-frequency pitch attitude gain would be necessary to provide acceptable attitude response to attitude command angles (θ_x in fig. 2) resulting from low-frequency position errors.

Position-loop crossover frequency was slightly larger for M_θ of 0 than for -3 and -5. The pilot did not suppress attitude disturbances as well at $M_\theta = 0$ and as a result they can be assumed to have caused increased position disturbances. Because hovering error did not increase at $M_\theta = 0$, the pilot apparently increased his response in the position loop. This is reflected in the slightly larger ω_{c_x} .

Pilot Adapted Parameters.—Pilot model adaptable parameters for the three values of M_θ are shown in the following tabulation:

M_θ	$M_{\delta e}$	K_{P_θ}	$K_{P_\theta} \cdot M_{\delta e}$	T_{L_θ}	K_{P_x}	T_{L_x}
0	0.403	20.0	8.1	0.54	0.035	0.36
-3	0.352	23.7	8.3	0.56	0.039	0.33
-5	0.327	21.0	6.9	0.50	0.048	0.33

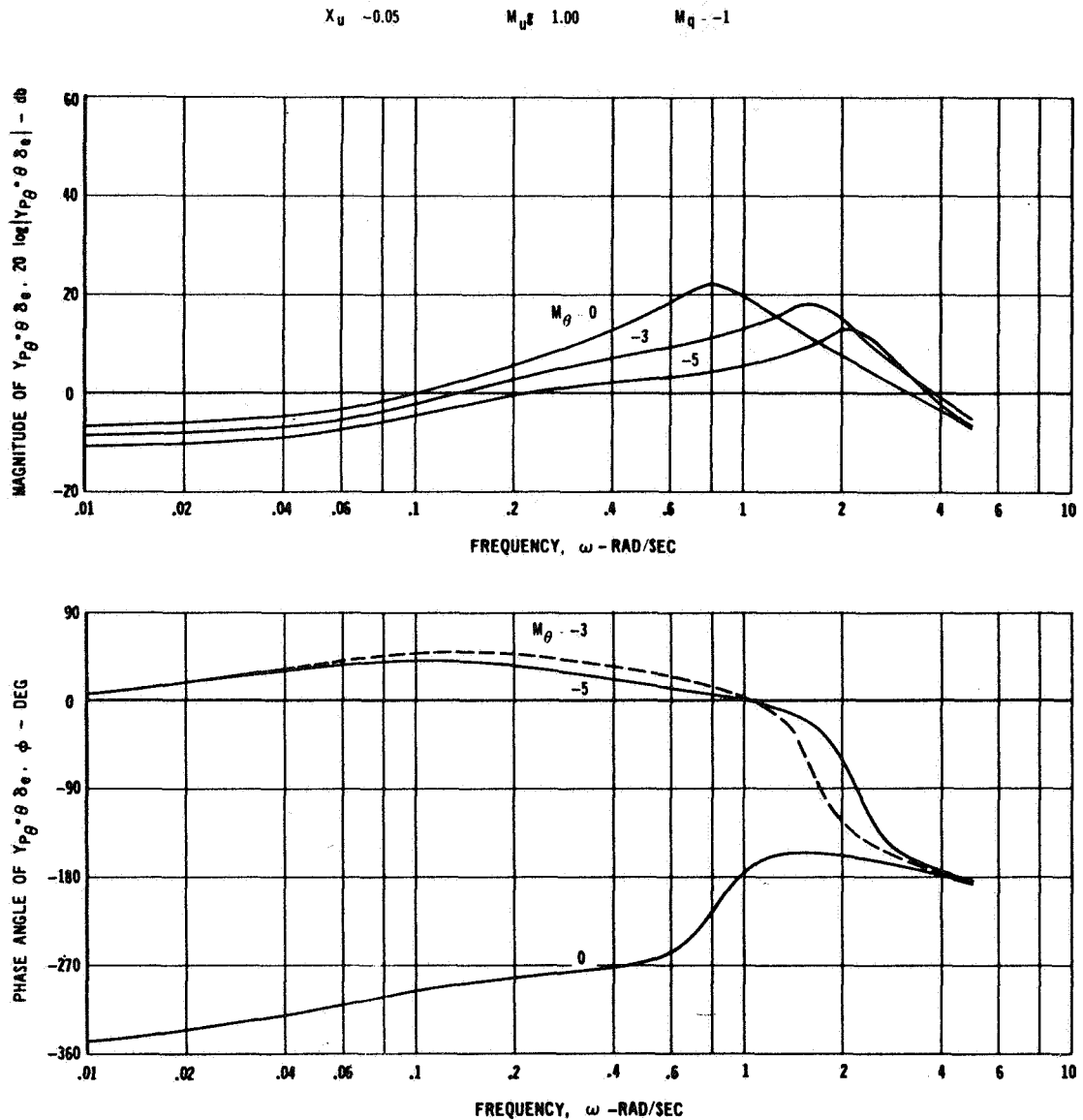


Figure 10.—Effect of pitch attitude stabilization on the frequency response of the transfer function relating pitch attitude to pitch error, $Y_{P_\theta}(s) \cdot \theta(s) / \delta_e(s)$.

The effective pilot gain, $K_{P_\theta} \cdot M_{\delta_e}$ increased slightly and then decreased with M_θ . Again, K_{P_θ} remained relatively constant while the pilot decreased M_{δ_e} . It was pointed out previously (fig. 9) that the addition of M_θ increased the magnitude of $\theta(s) / \delta_e(s)$ at frequencies greater than 1 to 1.5 rad/sec. Therefore, when the pilot adapts nearly the same level of gain for all values of M_θ , control stick sensitivity must be reduced as M_θ becomes more negative. Note, however, that the K_{P_θ} adapted here is only about half that generally adapted during the M_q and $M_{u\theta}$ studies. The reason for this difference is being investigated. As in the case of the M_q and $M_{u\theta}$ studies, it appears that gain $K_{P_\theta} \cdot M_{\delta_e}$ was adapted to provide the crossover frequencies, ω_{c_θ} , needed to control the high-frequency attitude disturbances.

Adapted pitch-loop lead remained essentially unchanged for all values of M_θ . It was also relatively high for all values of M_θ . Even though the aircraft dynamics became more

stable at $M_0 = -3$ (table 1), the pilot was required to extend ω_{c0} to the larger amplitudes of the high-frequency disturbances. As a result the lead requirements remained high because of the need to maintain closed-loop stability. This rationale also applies to the relatively high value of lead adapted for $M_0 = -5$.

The small changes in position-loop adapted parameters are consistent with the small changes in the pitch-loop parameters. These position-loop parameters did not have to change to compensate for the changes in pitch-loop parameters. Thus, they indicate the relatively constant nature of the position-loop closure.

REFERENCES

1. McRuer, D. T.; Graham, D.; Krendel, E. S.; and Reisener, W.: Human Pilot Dynamics in Compensatory Systems. AFFDL-TR-65-15, July 1965.
2. Stapleford, R. L.; McRuer, D. T.; and Magdaleno, R.: Pilot Describing Function Measurements in a Multi-Loop Task. NASA CR-542, Aug. 1966.
3. Bekey, G. A.: An Investigation of Sampled Data Models of the Human Operator in a Control System. ASD-TDR-62-36, Feb. 1962.
4. Miller, D. P.; and Vinje, E. W.: Fixed-Base Flight Simulator Studies of VTOL Aircraft Handling Qualities in Hovering and Low-Speed Flight. AFFDL-TR-67-152, prepared by United Aircraft Research Laboratories, Jan. 1968.
5. Seckel, E.; Traybar, J. J.; and Miller, G. E.: Longitudinal Handling Qualities for Hovering. Princeton University Rept. No. 594, Dec. 1961.
6. Friedman, G. R.: Helicopter Control: A Multi-Loop Manual Control System. M.I.T., Man-Vehicle Control Laboratory, Rept. No. MIT-67-2, June 1967.

8. Measurement of Pilot Describing Functions From Flight Test Data With an Example From Gemini X

*Rodney C. Wingrove and Frederick G. Edwards
Ames Research Center, NASA*

It is well known that there is an error in identifying the pilot describing function from routine flight test records because the pilot's output noise is correlated with the input error signal. This paper shows that this identification error can be reduced in the computer processing by shifting the input signal an amount equivalent to the pilot's time delay. This technique for reducing the identification error is analyzed with theory and is demonstrated with the identification of a simulated pilot model. This technique is also applied to flight test records obtained from the retrofire phase of the Gemini X mission.

INTRODUCTION

Pilot describing functions have usually been identified from records obtained in ground-based simulators and in flight tests wherein carefully controlled external forcing functions (e.g., sine waves) are used to excite the pilot-vehicle system. The pilot describing functions are measured by comparing the pilot's output and input signals with the known forcing function. This method minimizes those errors in identification due to any correlation of the input error signal with the pilot's output noise. Reference 1 contains a good review of this previous work and summarizes the measured pilot describing functions.

Most other methods for measuring pilot describing functions depend upon external disturbances (i.e., aerodynamic turbulence, propulsive disturbance, etc.) to excite the pilot-vehicle system. These are the open-loop methods that compute the pilot's describing function directly from the pilot's input and output signals. In reference 2, these methods have been reviewed and the expected errors in identification have been analyzed. It was shown that there is an identification error when using these open-loop methods because the pilot's output noise is transferred through the control loop, appearing as a component of his input error signal and, thus, is correlated with his input. It was further shown that if the pilot's noise was large, as compared with the external disturbance, then the identification error would probably be unacceptable.

During routine flight test operations, there are no carefully controlled forcing functions, and even the random external disturbance may be quite small. The purpose of this report is to present the development of a technique to reduce the error in identifying the pilot describing functions for these normal flight test operations. This technique involves shifting the time history of the input error signal during the computer processing. The input signal is

shifted an amount equivalent to the pilot's time delay. Although previous studies (e.g., refs. 3 to 5) have considered the use of a time shift in the measurement of pilot describing functions, it was apparently not observed that this time shift would strongly influence the errors in identification.

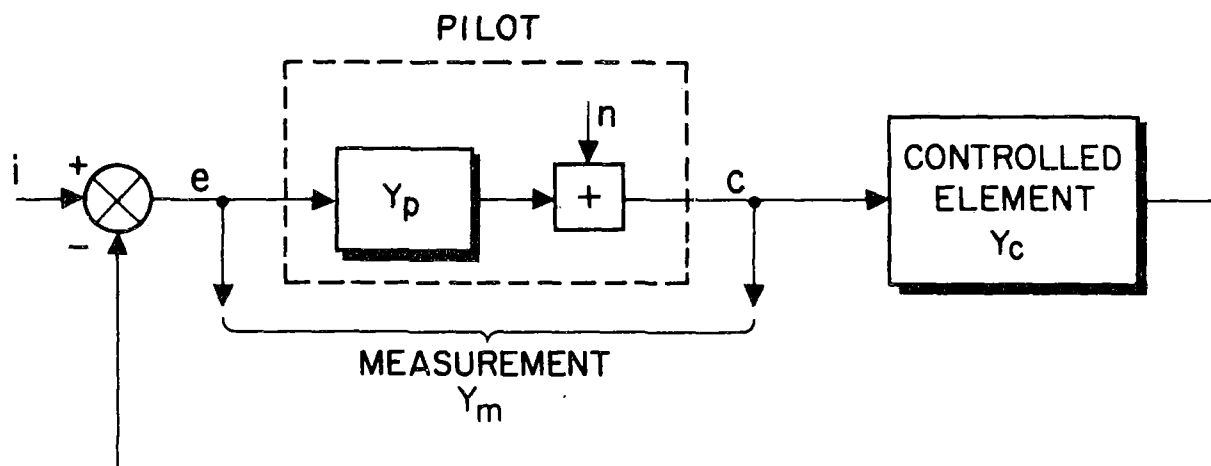
This paper presents in the appendix a theoretical analysis to show that this technique will reduce the identification error. The simulation and identification of a known system is included to illustrate the use of this technique. Also, results obtained from the retrofire phase of the Gemini X mission are presented to demonstrate the application of this technique to normal flight test records.

SYMBOLS

$c(t)$	controller deflection (output of pilot)
$e(t)$	error signal (input to pilot)
$i(t)$	external disturbance
$n(t)$	pilot noise
$R_{nn}(\tau)$	autocorrelation function of $n(t)$
t	time, sec
$Y_c(j\omega)$	controlled element
$Y_m(j\omega)$	measured transfer function
$Y_p(j\omega)$	pilot transfer function
$\hat{Y}_p(j\omega)$	estimated pilot transfer function
α	exponential decay factor, sec^{-1}
λ	time shift used during analysis, sec
τ_c	pure time delay in Y_c , sec
τ_p	pure time delay in Y_p , sec
$\Phi_{ee}(\omega)$	power spectrum of $e(t)$
$\Phi_{nn}(\omega)$	power spectrum of $n(t)$
$\Phi_{ec}(j\omega)$	cross-power spectrum of $e(t)$ and $c(t)$
$\Phi_{en}(j\omega)$	cross-power spectrum of $e(t)$ and $n(t)$
ω	frequency, rad/sec

IDENTIFICATION ERROR WITH OPEN-LOOP METHODS

The identification of pilot describing functions using open-loop measurement methods is illustrated in figure 1. To determine the unknown describing function, $Y_p(j\omega)$, there must be some signal within the control loop. This signal could be from the disturbance source, $i(t)$, or from noise introduced by the pilot, $n(t)$. Although $i(t)$ is shown in figure 1 as a time-varying command, it can also be considered (through block diagram reduction) to contain any other types of disturbances, such as aerodynamics, propulsion, etc., which are external to the pilot. The pilot is assumed to be in a compensatory tracking task trying to control his output $c(t)$ in such a manner as to keep the error signal $e(t)$ near zero.



CASE I $n \ll i$

$$Y_m(j\omega) = Y_p(j\omega)$$

CASE II $n \gg i$

$$Y_m(j\omega) = \frac{-1}{Y_c(j\omega)}$$

Figure 1. —Identification with open-loop methods.

Previous studies (e.g., ref. 2) have shown that there is a difference between the open-loop measurement, $Y_m(j\omega)$, and the actual describing function, $Y_p(j\omega)$. This difference, or identification error, can be analyzed using the following cross-spectral relationship:

$$Y_m(j\omega) = \frac{\Phi_{ec}(j\omega)}{\Phi_{ee}(\omega)} \quad (1)$$

where $\Phi_{ec}(j\omega)$ is the cross-power spectrum between $e(t)$ and $c(t)$ and $\Phi_{ee}(\omega)$ is the power density spectrum of $e(t)$. If basic relationships are used for the closed-loop system in figure 1, the cross-product can be written as the sum $\Phi_{ec}(j\omega) = Y_p(j\omega)\Phi_{ee}(\omega) + \Phi_{en}(j\omega)$. Substituting this into equation (1), we then obtain

$$Y_m(j\omega) = Y_p(j\omega) + \underbrace{\frac{\Phi_{en}(j\omega)}{\Phi_{ee}(\omega)}}_{\text{error}} \quad (2)$$

Thus, any correlation (i.e., $\hat{\Phi}_{en}(j\omega)$) between $e(t)$ and $n(t)$ is seen to contribute an error in identification. The size of this error is noted in the lower portion of figure 1. For case I, where $i(t)$ is much larger than $n(t)$, the measured describing function $Y_m(j\omega)$ will be near the true value $Y_p(j\omega)$. However, for case II, in which $n(t)$ is much larger than $i(t)$, the measured describing function will be very much in error and, in fact, reference 2 has shown that $Y_m(j\omega) = -1/Y_c(j\omega)$.¹ For routine flight test conditions, where $n(t)$ may be much larger than $i(t)$, it is necessary to find some means of reducing the error of case II. Such a technique will be outlined next.

USE OF A TIME SHIFT λ TO REDUCE IDENTIFICATION ERROR

We will now discuss a slight modification to the identification method of figure 1. This modification, which is illustrated in the upper portion of figure 2, involves the following steps in the computing process.

- (1) The input signal $e(t)$ is shifted with respect to $c(t)$ by an amount λ , where λ is equivalent to the pilot's time delay.
- (2) The transfer function $Y_m(j\omega)$ is determined from (1) using a measurement method in which $Y_m(j\omega)$ is constrained to be physically realizable (e.g. refs. 2 to 7).
- (3) The estimated transfer function is determined from the measured transfer function as

$$\hat{Y}_p(j\omega) = e^{-\lambda j\omega} Y_m(j\omega)$$

This technique is general and can be used with any measurement method in which $Y_m(j\omega)$ is constrained to be physically realizable (see appendix). For the purpose of this report, this technique will be illustrated with two representative methods: cross-correlation (refs. 2 and 6) and orthogonal filters (refs. 2 and 3).

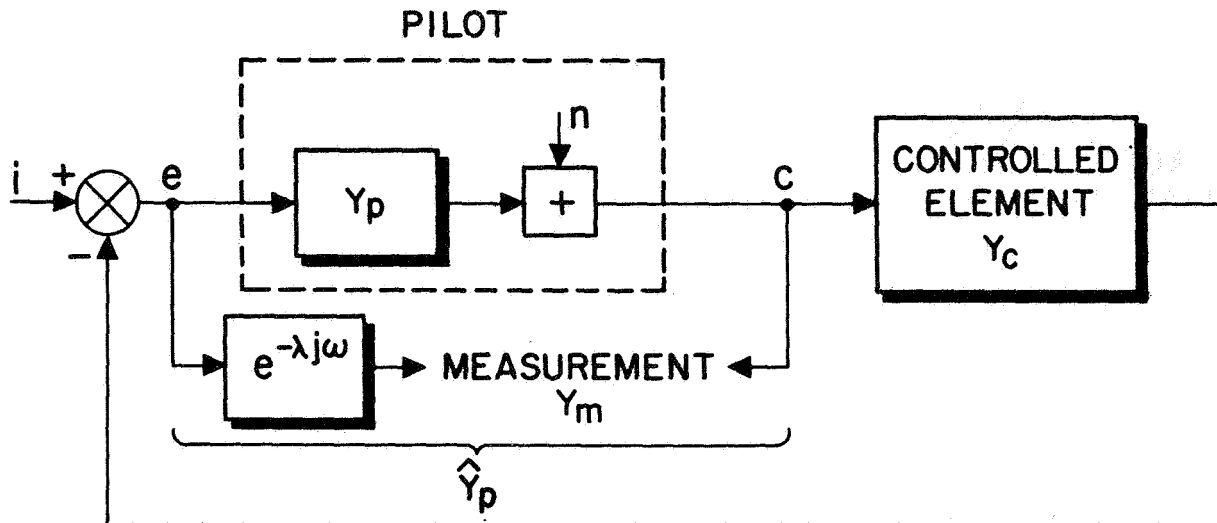
The theoretical development in the appendix analyzes the identification error to be expected with this type of computing process. The rest of this section will discuss the theoretical results in detail.

THEORETICAL RESULTS.—The theoretical development in the appendix assumes that time shift λ used in the computing process is less than or equal to pilot's time delay, τ_p . It has also been assumed in this development that $n(t) \gg i(t)$ in order to illustrate the maximum expected error in the identification of $Y_p(j\omega)$.

The appendix shows that the identification error depends upon the relationship between the pilot's impulse response function and his noise autocorrelation function, $R_{nn}(\tau)$. The theoretical results for two different conditions, case III and case IV, have been summarized in the lower portion of figure 2.

In case III, as shown in figure 3, $R_{nn}(\tau)$ is negligible for all values of τ greater than λ . For this case, which represents $n(t)$ near white noise, the theory shows that the identification error will be zero. This result appears to be significant and has many far-reaching ramifications. The most important point is that a system describing function, $Y_p(j\omega)$, can theoretically be identified with the system excited only by the internal noise, $n(t)$. The estimated describing function will be identical to the actual describing function, however, only if the special conditions noted in case III are met.

¹As shown in the appendix, this is modified if Y_c is nonminimum phase and the measurement has the constraint only to identify physically realizable systems.



CASE III $R_{nn}(\tau) = 0$ FOR $\tau > \lambda$

$$\hat{Y}_p(j\omega) = Y_p(j\omega)$$

CASE IV $R_{nn}(\tau) = Ke^{-\alpha|\tau|}$

$$\hat{Y}_p(j\omega) = Y_p(j\omega) - \underbrace{e^{-\alpha\lambda} \left(\frac{1}{Y_c(j\omega)} + Y_p(j\omega) \right)}_{\text{ERROR}} e^{-\lambda j\omega}$$

Figure 2. —The use of a time shift λ in identification; theory for $n \gg i$, $0 < \lambda \leq \tau_p$, and Y_c minimum phase.

A more general condition is noted under case IV. Here, $R_{nn}(\tau)$, the pilot's injected noise, described in terms of autocorrelation function, is assumed to take the form $R_{nn}(\tau) = Ke^{-\alpha|\tau|}$ which is narrow-band noise. This form agrees quite well with some experimental measurements of the pilot's remnant. For instance, this exponential form with $\alpha = 5 \text{ sec}^{-1}$ agrees with the measured $n(t)$ in reference 8. The identification error has been derived for case IV and is also shown in figure 2. Cases III and IV will now be illustrated by an example.

SIMULATION RESULTS.—To illustrate cases III and IV, the control loop shown in the upper left of figure 4 was simulated, and identification measurements were made on the known system. For this example, a representative pilot model and controlled element were as follows: $Y_p(j\omega) = 4e^{-0.3j\omega}$, $Y_c(j\omega) = 1/j\omega$. These measurements were made with no external disturbance, $i(t) = 0$, and the only excitation to the system dynamics was the internal noise source, $n(t)$. A time shift of $\lambda = 0.2$ second was used. Two forms of the noise spectrum were considered: an $n(t)$ with a spectrum that approximates white noise to illustrate case III, and $R_{nn}(\tau) = e^{-5|\tau|}$ to illustrate case IV.

For case III, where $n(t)$ is near white noise, the theory predicts that the measurement will identify the actual system; that is, $\hat{Y}_p(j\omega) = 4e^{-0.3j\omega}$. Figure 4 presents the experimental

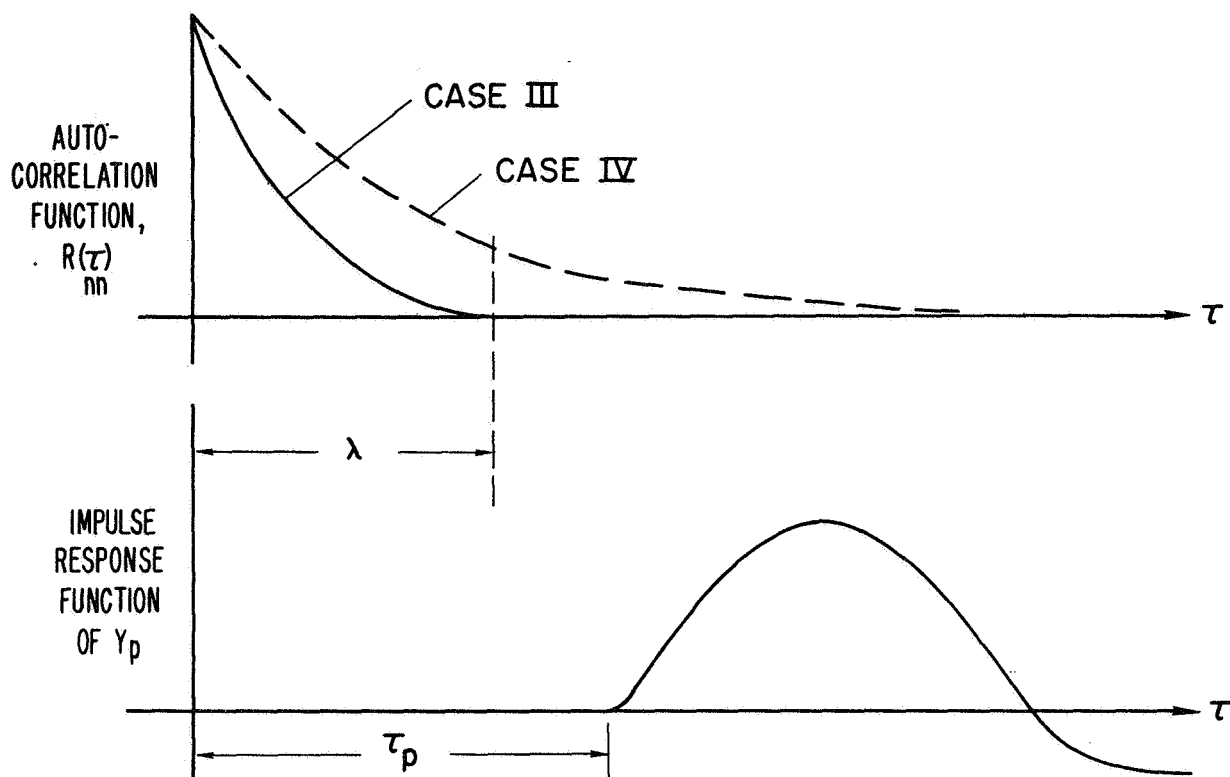


Figure 3. —Relationships between $R_{nn}(\tau)$ and the system response function.

results. The measured amplitude of $\hat{Y}_p(j\omega)$ is only ± 0.5 dB about the actual value for frequencies to about 9 rad/sec and the phase angle varies $\pm 0.5^\circ$ about the actual value. These differences are believed to be within the experimental accuracies of the simulation. The results for case III seem to substantiate the theoretical conclusion that it is possible to identify the describing function of a system that is excited by noise $n(t)$ introduced internally within the system.

For case IV, where $R_{nn}(\tau) = e^{-5|\tau|}$, theory predicts an error in the identification; that is,

$$\hat{Y}_p(j\omega) = \underbrace{4e^{-0.3j\omega}}_{Y_p(j\omega)} - \underbrace{0.37(j\omega + 4e^{-0.3j\omega})e^{-0.2j\omega}}_{\text{error}} \quad (3)$$

The experimental data for case IV, in figure 4, are again near the value predicted by the theory. From this figure, the estimated value of the amplitude $|\hat{Y}_p(j\omega)|$ is seen to produce too low a value (about 4 dB below the true value) at the lower frequencies and tends to give the appearance of lead (slope = 20 dB/decade) at the higher frequencies. Actually, the measurement is tending toward $-1/Y_c(j\omega)$ as predicted by the theory. The phase angle, however, agrees quite well with the true value.

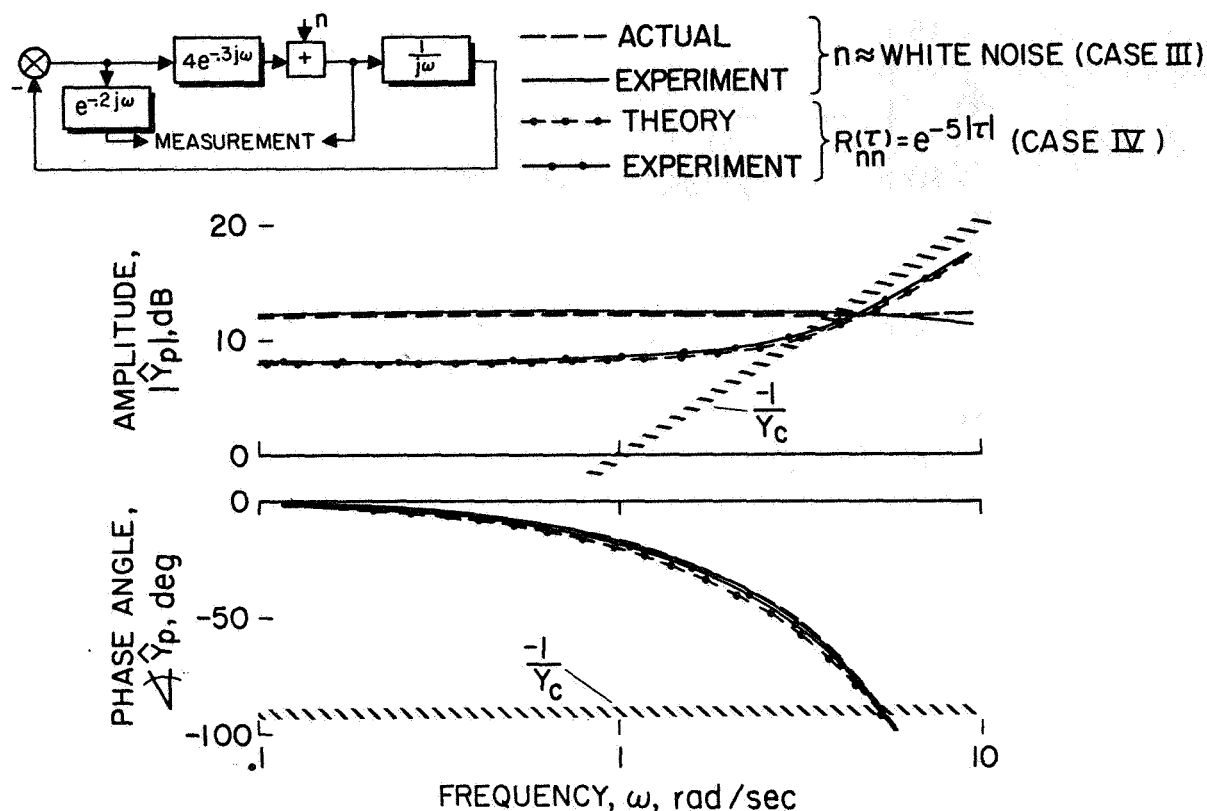


Figure 4. —Example using only internal disturbance.

EFFECT OF TIME SHIFT λ . —Experimental data are presented in figure 5 for several values of λ . These results are for one value of frequency, $\omega = 1 \text{ rad/sec}$, and for the control system and spectrum of the excitation noise source, $R_{nn}(\tau) = e^{-5|\tau|}$, as used in the previous example.

In this figure the experimental data are compared with the value for the actual system ($|Y_p| \approx 12 \text{ dB}$ and $\angle Y_p \approx -18^\circ$). This comparison illustrates that the identification error decreases, as predicted by theory, out to a value of $\lambda = 0.3 \text{ second}$, which is equal to τ_p . (The theory, as noted previously, is only valid for $\lambda \leq \tau_p$.) The error in $|\hat{Y}_p(j\omega)|$ is seen to increase for values of λ much larger than τ_p . This is to be expected because $Y_m(j\omega)$ cannot properly model $Y_p(j\omega)$ when $\lambda > \tau_p$. These data indicate that the minimum error in $|\hat{Y}_p(j\omega)|$ occurs near a value of $\lambda \approx \tau_p$. The error in the phase angle $\hat{Y}_p(j\omega)$ is also seen to be very small at $\lambda \approx \tau_p$.

As shown by this example, the value of λ should be near the pilot's time delay τ_p to minimize the error in identification. The pilot's time delay may be approximately known in some situations (i.e., ref. 1) but, in general, its value will be unknown and will depend upon the particular piloting task. One method of estimating τ_p (thus selecting λ) will be illustrated later in the report.²

²In references 3 to 5, the value for λ which gave the best linear fit between $e(t)$ and $c(t)$ was used. This leads to an erroneous result because if $n(t) \gg i(t)$, then the best fit is obtained when $\lambda = 0$ and in this case $\hat{Y}_p(j\omega) = -1Y_c(j\omega)$.

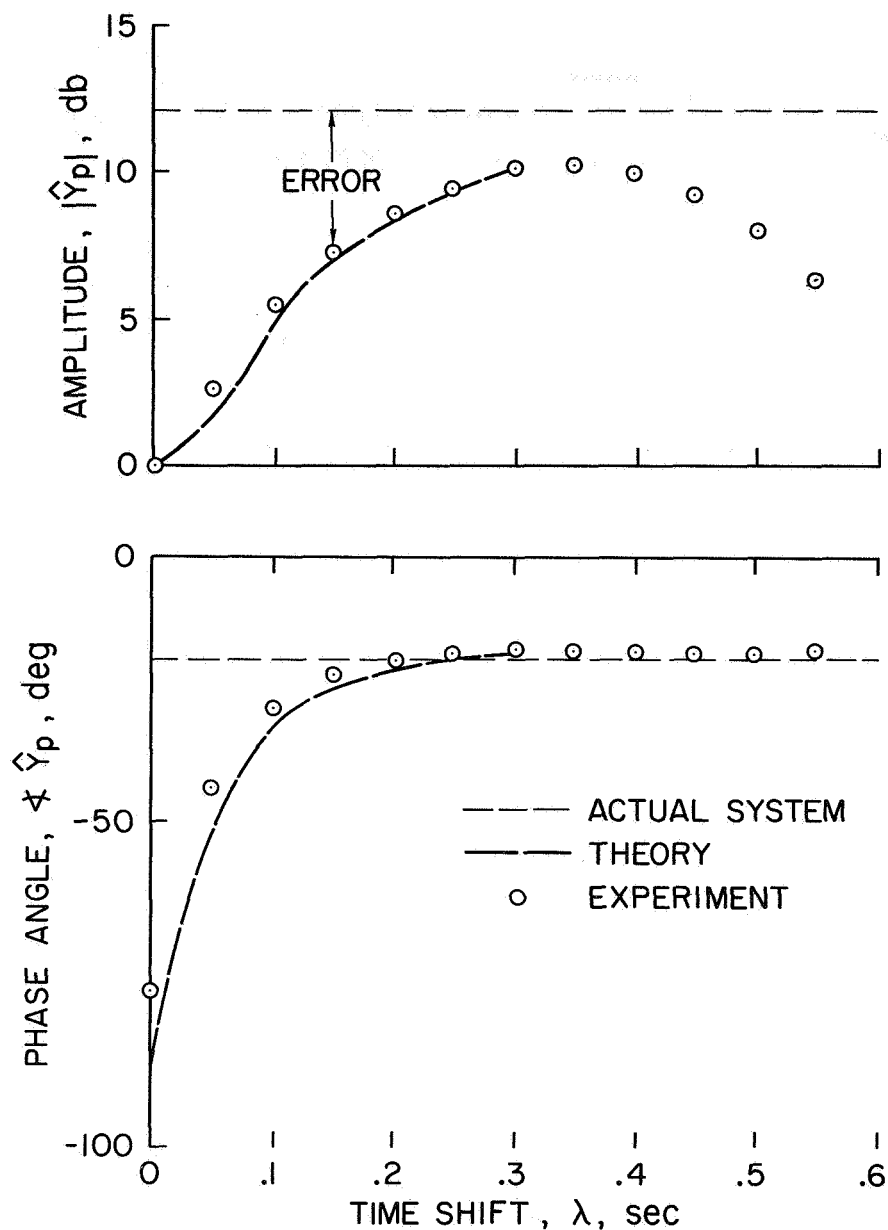


Figure 5.—The effect of λ on the identification error; $\omega = 1$ rad/sec.

FLIGHT TEST RESULTS FROM GEMINI X

Flight data from Gemini X was analyzed to illustrate the application of this identification technique. In analyzing flight test data, it is important to select a section of the record that contains disturbances external to the pilot. As we noted earlier, external disturbances will tend to reduce the error in identification. The retrofire maneuver is an example when external disturbances were introduced because of the unsymmetric ripple firing of the four retrorockets.

During retrofire, the pilot is controlling the attitude about each of the three axes. There is no control coupling between these axes and the pilot appears to treat them as three

separate tasks. Of the three axes, the control about the yaw axis, shown in figure 6, was found to contain the best consistent correlation between attitude deviations, $e(t)$ and control stick deflections, $c(t)$. These data for the yaw axis control will be used to illustrate the measurement of the pilot's describing function during the retrofire of the Gemini spacecraft.

The Bode plots of the pilot describing function obtained by the cross-correlation method for the data of figure 6 are presented in figure 7. Curves of the magnitude $|\hat{Y}_p(j\omega)|$ and phase angle $\angle \hat{Y}_p(j\omega)$ are presented as a function of frequency for two values of λ : $\lambda=0$ and $\lambda=0.7$ second. Also shown by the cross-hatched line is the Bode plot for the approximation to $-1/Y_c(j\omega)$ (this is an approximation because Y_c is not linear). The significance of this line was noted previously. The theory predicts that for $\lambda=0$, the estimated describing function $\hat{Y}_p(j\omega)$ will tend toward $-1/Y_c(j\omega)$ as illustrated in figure 7. However, for this flight situation, $\hat{Y}_p(j\omega)$ does not coincide exactly with $-1/Y_c(j\omega)$ because of external disturbances due to the firing of the retrorockets and the jet-control system.

For $\lambda=0.7$ second, the estimated describing function $\hat{Y}_p(j\omega)$ tends away from the curve of $-1/Y_c(j\omega)$. Any value of λ from about $\lambda=0.3$ to 0.9 second resulted in approximately the same Bode plot as shown for $\lambda=0.7$ second. This estimated describing function can be approximated by a transfer function with a constant gain and a time delay, $Y_p(j\omega) \approx Ke^{-\tau_p j\omega}$.

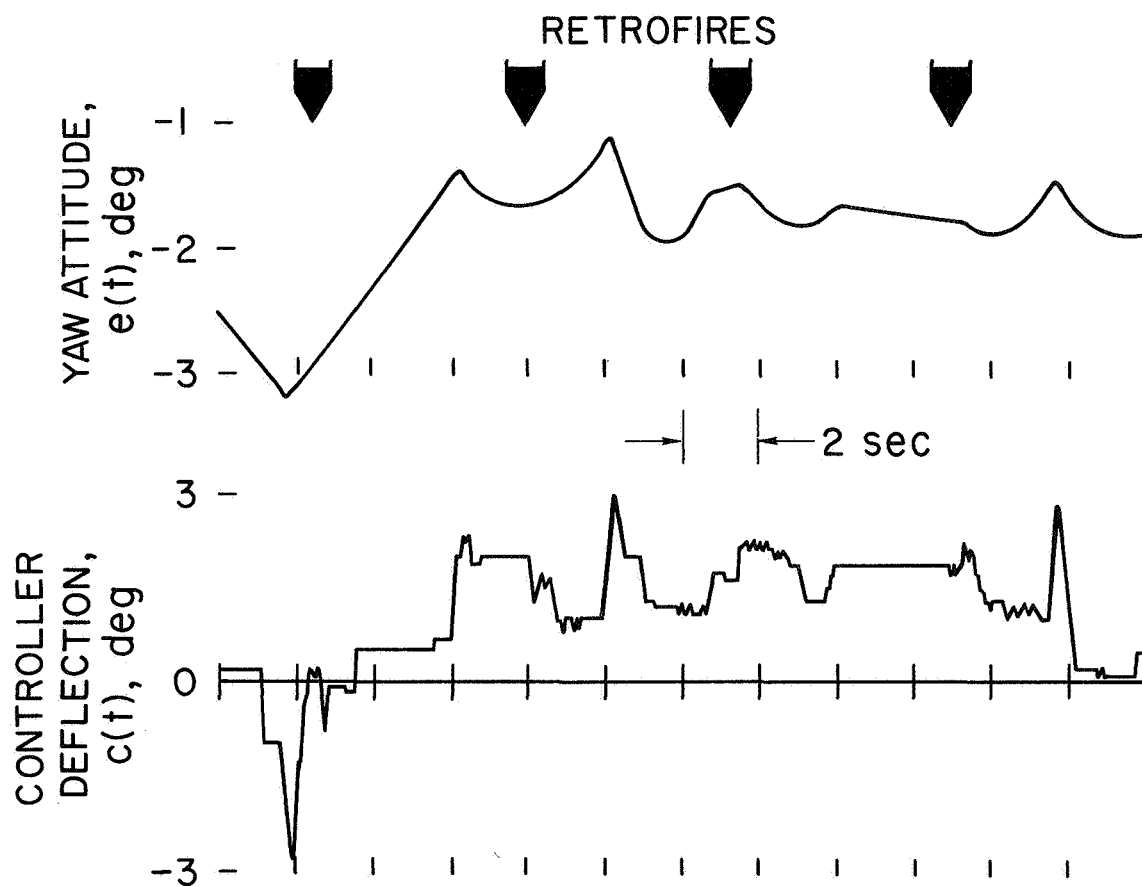


Figure 6.—Time history of yaw control task during retrofire.

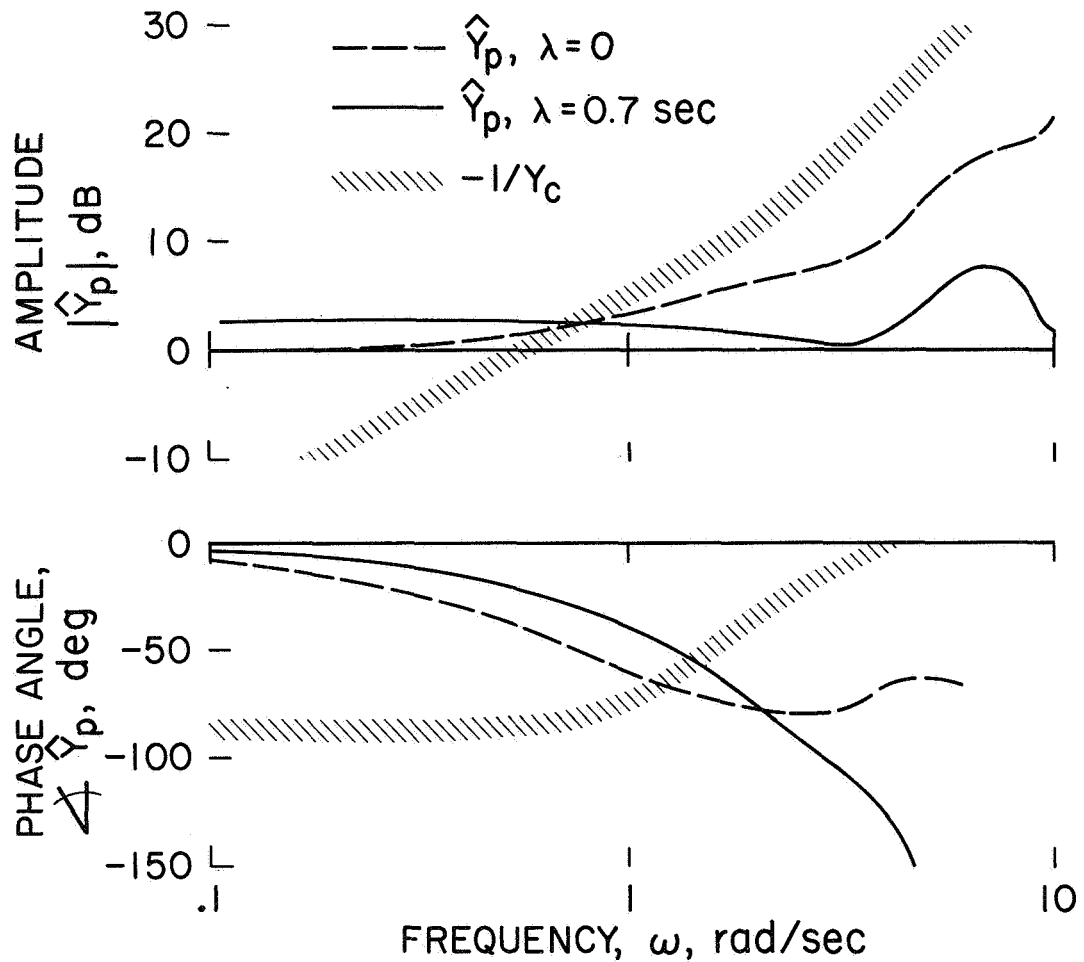


Figure 7.—The effect of λ on identification. Effect of time delay on measurement. Yaw control during retrofire.

As noted previously, the value of λ that will minimize the identification error is dependent upon the value of the pilot's time delay τ_p . For these data, the following procedure was used to determine τ_p and, thus, select λ .

- (1) Compute the Bode plot for a given value of λ
- (2) Select the transfer function that best fits the Bode plot (i.e., $Y_p(j\omega) = Ke^{-\tau_p j\omega}$)
- (3) Note the value of τ_p from (2)
- (4) Repeat steps (1) through (3) until a value of $\lambda \approx \tau_p$ is obtained

This procedure is illustrated in figure 8 for the yaw axis data. The estimated τ_p 's from fairing through the Bode plots are presented as a function of λ . It is seen that λ is equal to the estimated time delay, τ_p , at $\lambda \approx 0.7$ second. Therefore, $\lambda = 0.7$ second was selected for use in this identification analysis.

The estimated describing functions that have been presented so far in this paper were determined using the cross-correlation measurement method. To illustrate that the technique of using λ is not restricted to any one method, additional results for the orthogonal filters method are presented in figure 9. This figure compares the Bode plots obtained with $\lambda = 0.7$ sec, using both the orthogonal filters and cross-correlation methods. We see that,

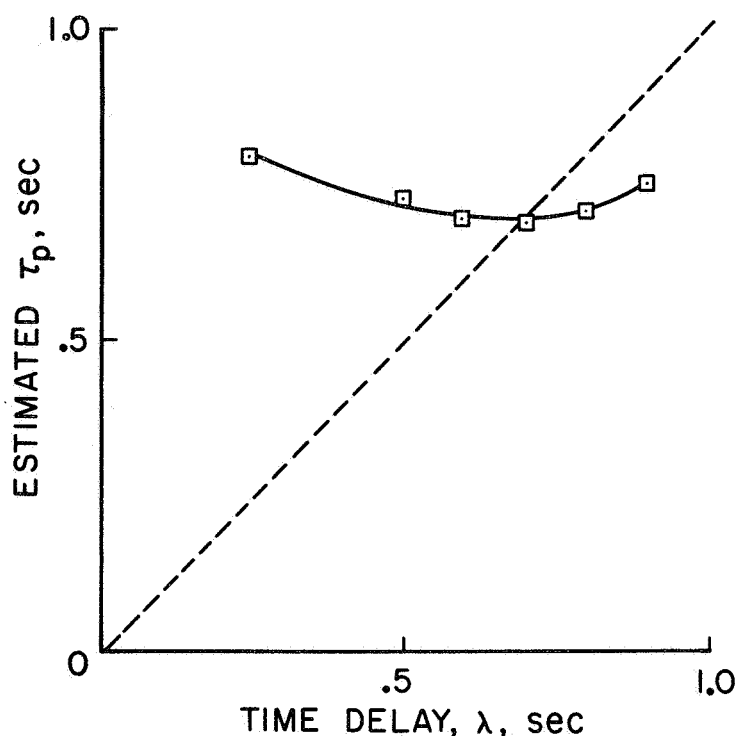


Figure 8.—Comparison of estimated τ_p with λ .
Yaw control during retrofire.

for frequencies ≤ 2 rad/sec, the estimated describing functions are in good agreement. The major differences appear at the high frequencies. However, for the flight records used in this analysis, there is very little input power at frequencies above about 2 rad/sec so the data shown at frequencies above this value probably have little significance.

The describing functions in figure 9 represent a constant gain system with a time delay, τ_p , of about 0.7 second. This result, although not directly comparable to the results from previous studies, does appear reasonable. For instance, with a rate command system as used in this control task, reference 1 has shown that the pilot describing function will be essentially a constant gain system with a pure time delay. The value τ_p for the three-axis flight data is higher than the value from the single-axis data in reference 1. However, other studies, such as in references 9 and 10, have also shown higher values of τ_p when the pilot is involved in the complete task of monitoring the instrument panel and controlling about three separate axes.

EFFECT OF TIME LAGS IN THE TOTAL CONTROL LOOP

Up to this point, we have assumed that $Y_c(j\omega)$ is minimum phase; for instance, we have not included the effect of any transport lags in $Y_c(j\omega)$. As shown in the appendix, any time delay, τ_c , in $Y_c(j\omega)$ will decrease the measurement error in $\hat{Y}_p(j\omega)$. In particular, the measurement error will theoretically be zero if

$$R_{nn}(\tau) = 0 \text{ for } \tau > \lambda + \tau_c$$

where

$$\lambda \leq \tau_p$$

This means that, in general, the identification error can be made small if the autocorrelation function of the internal noise is negligible when τ is greater than the sum of all transport lags through the total control loop, $\tau_c + \tau_p$.

This approach to system identification, as developed in this paper, indicates that the internal noise, $n(t)$, need not be a hinderance to identification; but, rather, it will aid in the identification and analysis of feedback control systems if the conditions stated above are met. This is an important point that may also have application in many other fields (i.e., biology, economics, chemical processes) where measurements can only be made with the noise introduced internally within the system to be measured.

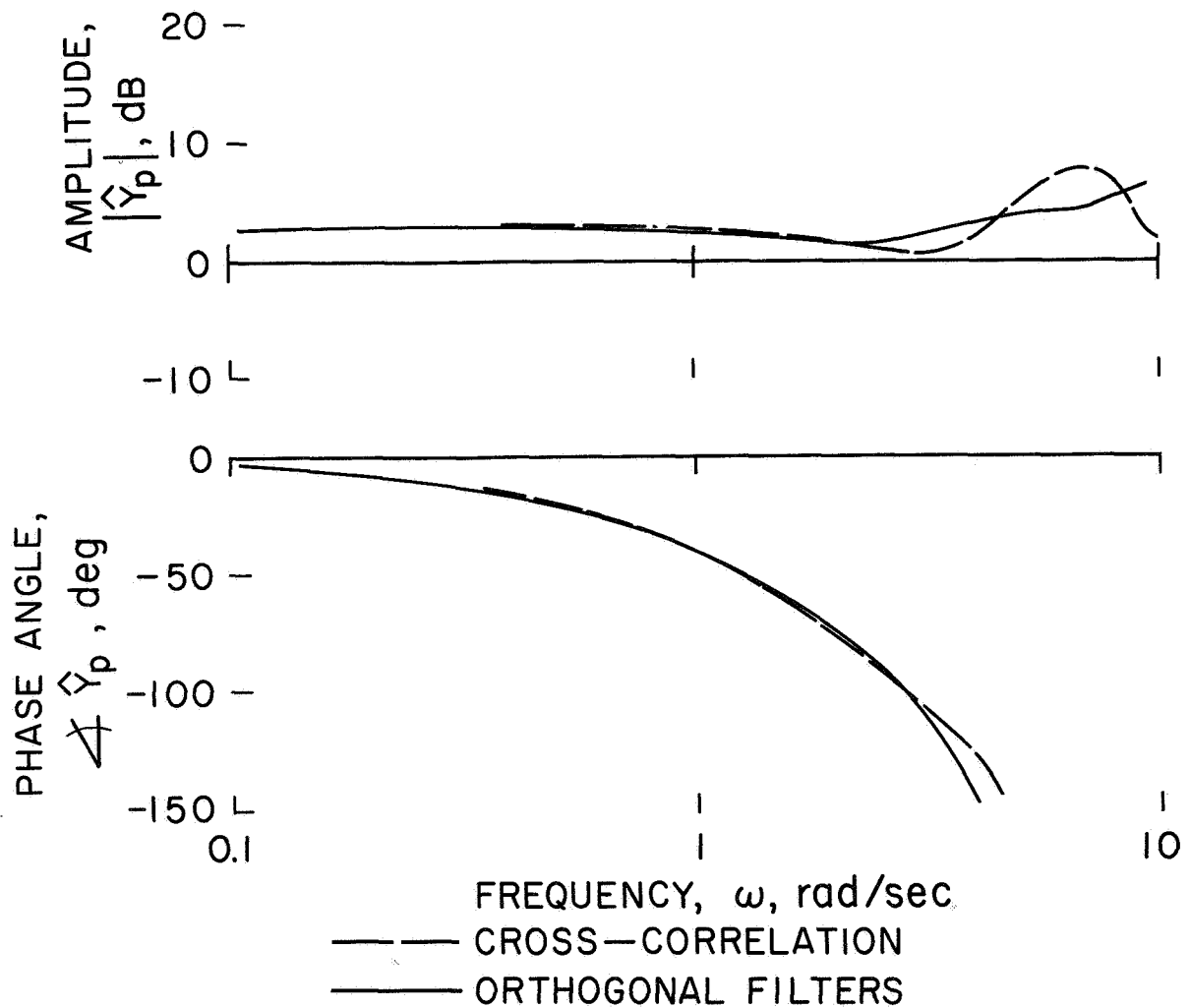


Figure 9. —Comparison for two measurement methods; $\lambda = 0.7$ sec.

CONCLUDING REMARKS

This paper has shown that in measuring pilot describing functions, the identification error due to the correlation of the input error signal with the pilot's output noise can be reduced by shifting the input data during the computer analysis. The value for this time shift should be near the time delay of the pilot. It is shown that this identification error can be made small if the autocorrelation function, $R_{nn}(\tau)$, of the internal noise source is negligible for τ greater than the sum of all transport lags through the control loop. This means that if these conditions are met, it is possible to measure the describing function of a system with feedback using only its own internal noise source for excitation.

Representative data selected from the retrofire portions of the Gemini X flight was analyzed using the technique outlined in this paper. These results demonstrate the feasibility of measuring the pilot's describing function from normal flight test records.

APPENDIX

ANALYSIS OF THE ERROR IN IDENTIFICATION.—The subject of this appendix is the open-loop identification error due to the correlation of $e(t)$ with $n(t)$ when the measurement method is constrained to be physically realizable.³ This constraint of physical realizability follows the usual definition and means that the impulse response function is zero for negative values of time. In order to mathematically represent the measured transfer function $Y_m(j\omega)$ with this constraint, we can utilize the relationship used with the Wiener-Hopf equations (ref. 11). Using this relationship for physically realizable systems, equation (1) is written.

$$Y_m(j\omega) = \frac{1}{\Phi_{ee}^+(j\omega)} \left[\frac{\Phi_{ec}(j\omega)}{\Phi_{ee}^-(j\omega)} \right]_+ \quad (A-1)$$

where

$$\begin{aligned} \Phi_{ee}(\omega) &= \Phi_{ee}^+(j\omega) \Phi_{ee}^-(j\omega) \\ \Phi_{ee}^+(j\omega) &\text{ has poles or zeros only in the LHP} \\ \Phi_{ee}^-(j\omega) &\text{ has poles or zeros only in the RHP} \\ [\]_+ &\text{ has poles only in the LHP} \end{aligned}$$

This follows the usual form which implies that the direct transform of a time function that is stable and is zero for negative time will have all its poles in the left-half plane (LHP).

³This constraint is inherent in the computer programming techniques for most time-domain measurement methods such as cross-correlation (e.g., refs. 2 and 6), orthogonal filters (e.g., refs. 2, 3, and 5) and parameter trackers (e.g., refs. 2, 4, and 9). Most frequency domain measuring methods using cross-spectral computing programs (ref. 1) usually do not contain this constraint. However, such a constraint could probably be incorporated.

Now we introduce the time shift λ as illustrated in figure 2 and define the shifted data as $e'(t) = e(t-\lambda)$. Because $\hat{\Phi}_{e'c}(j\omega) = e^{\lambda j\omega} \hat{\Phi}_{ec}(j\omega)$ and $\hat{\Phi}_{e'e'}(\omega) = \hat{\Phi}_{ee}(\omega)$, we can write the measured transfer function as

$$Y_m(j\omega) = \frac{1}{\hat{\Phi}_{ee}^+(j\omega)} \left[\frac{e^{\lambda j\omega} \hat{\Phi}_{ec}(j\omega)}{\hat{\Phi}_{ee}^-(j\omega)} \right]_+$$

Then, as shown in figure 2, we define the estimated describing function in terms of measured transfer function, $\hat{Y}_p(j\omega) = e^{-\lambda j\omega} Y_m(j\omega)$, and arrive at

$$\hat{Y}_p(j\omega) = \frac{e^{-\lambda j\omega}}{\hat{\Phi}_{ee}^+(j\omega)} \left[\frac{e^{\lambda j\omega} \hat{\Phi}_{ec}(j\omega)}{\hat{\Phi}_{ee}^-(j\omega)} \right]_+ \quad (A-2)$$

Introducing the individual terms for $\hat{\Phi}_{ec}(j\omega)$ (see eq. (2)), we have

$$\hat{Y}_p(j\omega) = \frac{e^{-\lambda j\omega}}{\hat{\Phi}_{ee}^+(j\omega)} \left[e^{\lambda j\omega} Y_p(j\omega) \hat{\Phi}_{ee}^+(j\omega) \right]_+ + \frac{e^{-\lambda j\omega}}{\hat{\Phi}_{ee}^+(j\omega)} \left[\frac{e^{\lambda j\omega} \hat{\Phi}_{en}(j\omega)}{\hat{\Phi}_{ee}^-(j\omega)} \right]_+ \quad (A-3)$$

The impulse response of the pilot describing function $Y_p(j\omega)$ is zero for time less than τ_p and so, as long as λ is less than or equal to τ_p , the term $e^{\lambda j\omega} Y_p(j\omega)$ is physically realizable. Simplifying equation (A-3) with this assumption, we obtain

$$\hat{Y}_p(j\omega) = Y_p(j\omega) + \frac{e^{-\lambda j\omega}}{\hat{\Phi}_{ee}^+(j\omega)} \left[\frac{e^{\lambda j\omega} \hat{\Phi}_{en}(j\omega)}{\hat{\Phi}_{ee}^-(j\omega)} \right]_+ \quad (A-4)$$

The term $\hat{\Phi}_{ee}(\omega)$ is made up of contributions from two sources; $i(t)$ and $n(t)$. The maximum error can be determined by assuming $i(t)=0$ (ref. 2). With this assumption and using basic closed-loop relationships, let us define

$$\hat{\Phi}_{ee}^+(j\omega) = \frac{-Y_c(j\omega) \hat{\Phi}_{nn}^+(j\omega)}{1 + Y_p Y_c(j\omega)} \quad (A-5)$$

$$\hat{\Phi}_{ee}^-(j\omega) = \frac{-Y_c(-j\omega) \hat{\Phi}_{nn}^-(j\omega)}{1 + Y_p Y_c(-j\omega)} \quad (A-6)$$

$$\hat{\Phi}_{nn}(j\omega) = \frac{-Y_c(-j\omega)\hat{\Phi}_{nn}^+(j\omega)\hat{\Phi}_{nn}^-(j\omega)}{1 + Y_p Y_c(-j\omega)} \quad (A-7)$$

These definitions assume that $Y_c(j\omega)$ is minimum phase (i.e., contains no time delay or zeros in RHP). A case in which $Y_c(j\omega)$ is nonminimum phase will be illustrated at the end of this appendix.

Y_c Minimum Phase.—Using the foregoing assumptions, which cover a broad variety of piloted control situations, we arrive at

$$\hat{Y}_p(j\omega) = Y_p(j\omega) - \frac{[e^{\lambda j\omega}\hat{\Phi}_{nn}^+(j\omega)]_+}{e^{\lambda j\omega}\hat{\Phi}_{nn}^+(j\omega)} \left[\frac{1}{Y_c(j\omega)} + Y_p(j\omega) \right] \quad (A-8)$$

where $[e^{\lambda j\omega}\hat{\Phi}_{nn}^+(j\omega)]_+$ is evaluated as $\mathcal{Z}u(t)R_{nn}(t + \lambda)$, in which u is the unit step function, and $R_{nn}(\tau)$ is the autocorrelation function of the noise, $n(t)$. The above means that the Fourier transform is only evaluated for τ greater than λ .

Equation (A-8) indicates those conditions under which the measurement error will be small. For instance, note that if λ is positive and if $n(t)$ is white noise ($R_{nn}(\tau)$ is an impulse at $\tau = 0$), then $[e^{\lambda j\omega}\hat{\Phi}_{nn}^+(j\omega)]_+$ is zero and there will be no error in identification. The general requirement for the measurement error to be zero is that $R_{nn}(\tau) = 0$ for $\tau > \lambda$.

Let us further look at a more general form for $n(t)$ and let $R_{nn}(\tau) = Ke^{-\alpha|\tau|}$ and we obtain

$$\hat{Y}_p(j\omega) = Y_p(j\omega) - e^{-\alpha\lambda} \left[\frac{1}{Y_c(j\omega)} + Y_p(j\omega) \right] e^{-\lambda j\omega} \quad (A-9)$$

This shows that the error term on the right side of the equation is a function of the magnitude of the constant factor $e^{-\alpha\lambda}$. As λ increases, and if α is large (near white noise), then $\hat{Y}_p(j\omega) \approx Y_p(j\omega)$. Conversely, if $\lambda = 0$ then the result is identical to that shown in reference 2; $\hat{Y}_p(j\omega) = -1/Y_c(j\omega)$.

Y_c Nonminimum Phase.—Let us define the nonminimum phase terms as: τ_c to represent any pure time delay in $Y_c(j\omega)$; $N_c^-(j\omega)$ to represent any RHP zeros in $Y_c(j\omega)$. Then, by including these terms in equations (A-5) to (A-7), the estimated describing function becomes

$$\hat{Y}_p(j\omega) = Y_p(j\omega) - \frac{[e^{(\lambda+\tau_c)j\omega}\hat{\Phi}_{nn}^+(j\omega)N_c^-(-j\omega)/N_c^-(j\omega)]_+}{e^{(\lambda+\tau_c)j\omega}\hat{\Phi}_{nn}^+(j\omega)N_c^-(-j\omega)/N_c^-(j\omega)} \left[\frac{1}{Y_c(j\omega)} + Y_p(j\omega) \right] \quad (A-10)$$

It is interesting to note that in this case if $R_{nn}(\tau) = 0$ for $\tau > \tau_c$, then $Y_p(j\omega)$ does not have to have a time delay (and no time delay, λ , is required in the analysis) in order to have zero measurement error.

REFERENCES

1. McRuer, Duane; Graham, Dunstan; Krendel, Ezra; and Reisener, William, Jr.: Human Pilot Dynamics in Compensatory Systems. Tech. Rept. AFFDL TR-65-15, USAF, July 1965.
2. Elkind, Jerome I.: Further Studies of Multiple Regression Analysis of Human Pilot Dynamic Response; A Comparison of Analysis Techniques and Evaluation of Time-Varying Measurements. ASD-TDR-63-618, Mar. 1964.
3. Elkind, Jerome I.; Starr, Edward A.; Green, David M.; and Darley, D. Lucille: Evaluation of a Technique for Determining Time-Invariant and Time-Variant Dynamic Characteristics of Human Pilots. NASA TN D-1897, 1963.
4. Todosiev, E. P.; Rose, R. E.; Bekey, G. A.; and Williams, H. L.: Human Tracking Performance in Uncoupled and Coupled Two-Axis Systems. NASA CR-532, 1966.
5. Wierwille, Walter W.; and Gagne, Gilbert A.: A Theory for the Optimal Deterministic Characterization of the Time-Varying Dynamics of the Human Operator. NASA CR-170, Feb. 1965.
6. Taylor, Lawrence W.: A Comparison of Human Response Modeling in the Time and Frequency Domains. Third Annual NASA-University Conference on Manual Control. NASA SP-144, 1967, pp. 137-153.
7. Goodman, T. P.; and Reswich, J. B.: Determination of System Characteristics from Normal Operating Records. Trans. ASME, vol. 7, Feb. 1956, pp 258-271.
8. McDonnell, J. D.; and Jex, H. R.: A Critical Tracking Task for Man-Machine Research Related to the Operators Effective Time Delay. NASA CR-674, 1966.
9. Adams, James J.; Bergeron, Hugh P.; and Hert, George J.: Human Transfer Functions in Multi-Axis and Multi-Loop Control Systems. NASA TN D-3305, Apr. 1966.
10. Clement, W. F.; Jex, H. R.; and Graham, D.: Application of a System Analysis Theory for Manual Control Displays to Aircraft Instrument Landing. Presented at 4th Annual NASA-University Conference on Manual Control, Univ. of Michigan, March 1968.
11. Newton, George C.; Gould, Leonard A.; and Kaiser, James F.: Analytical Design of Linear Feedback Controls. John Wiley & Sons, Inc., 1957.

9. Models for Steering Control of Motor Vehicles

*David H. Weir and Duane T. McRuer
Systems Technology, Inc.*

Driver control of an automobile is analyzed using feedback models for the system that consists of the driver, the vehicle, and the roadway environment. Three alternative models are evolved that satisfy the guidance and control requirements of the driver/vehicle system. They are: time-advanced lateral position, path angle plus lateral position, and heading angle plus lateral position. Although there may be other alternatives, these provide good performance, insensitivity to variations in the driver's dynamic adaptation, good predicted subjective opinion rating, etc.; and they are compatible with perceptual data from driving experiments. The resultant models give the highway and vehicle designer a tool that can be useful in analyzing the system and in optimizing its parameters for the safest and most efficient operation.

INTRODUCTION

Some form of steering control is required from the driver to make a vehicle perform a maneuver or follow the desired path. To understand driver control quantitatively it is necessary to consider the actions of random and deterministic inputs on a dynamic system comprising a vehicle whose equations of motion are known and a driver whose dynamic response under various conditions can be estimated.

Two things are needed to accomplish the driving task:

- (1) A desired path (command input) that has adequate tolerances to offset variations in decision and judgmental processes
- (2) A guidance and control system structure to execute the command inputs with reasonable ease and precision

Recent research (refs. 1, 2, and 3) has centered on the guidance and control area; this has included the derivation, validation, and exercise of appropriate operational models for the driver/vehicle/roadway system.

This paper is devoted to studies of steering or directional control. These are important for many reasons, not the least being that many hazardous driving situations involve steering difficulties and that this type of control is poorly understood for any phase of driving. Throttle and braking (longitudinal) control by the driver is also important, but it has been studied more extensively (e.g., refs. 4 to 10) and it is relatively well understood.

The paper begins with a short overview of the system structure to place the component elements in an overall system context. Then the components, that is, the dynamics of the vehicle and of the driver, are summarized. With the elements defined in some detail, the structure of the driver/vehicle closed-loop system is then reexamined thoroughly, considering both single-loop and multiloop forms. Finally, some implications of the theory are presented.

OVERVIEW OF THE MODEL STRUCTURE

The driver, the vehicle, and the roadway environment are the three essential parts of the closed-loop structure for a single vehicle element. The general topology of this structure for steering control is presented in figure 1. It illustrates the possible types of driver response blocks, how the driver interacts with the vehicle, and how the driver/vehicle system interacts with the roadway environment.

Three general levels of control structure are shown for the driver in figure 1. They correspond to three levels of driver behavior, and consequent system performance, and are elements of the Successive Organization of Perception in reference 11. The levels are:

(1) Precognitive which involves executing a learned maneuver in an open-loop way. It is typified by the Internally Generated Maneuver Command block. The command comes from within the driver after being triggered by some pattern or stimulus in the visual and/or proprioceptive field. Examples might include turning into one's driveway, or portions of an overtaking and passing maneuver such as the initial pullout.

(2) Pursuit which takes advantage of a knowledge of the system input to structure a driver feedforward which improves performance (ref. 12). It is shown in figure 1 as the

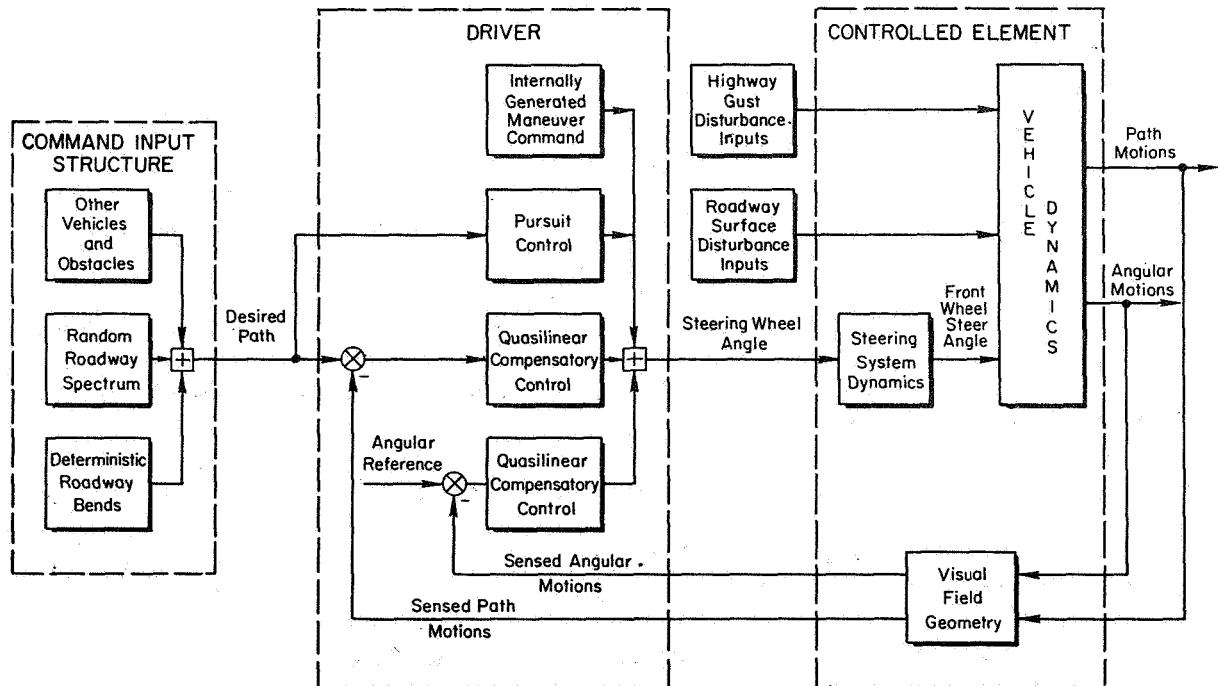


Figure 1. —Topology of the closed-loop control structure.

Pursuit Control block. The essence of pursuit behavior is the combined open-loop/closed-loop characteristic. The open-loop feedforward element provides a driver output that causes the vehicle output to very nearly duplicate the command input, while the closed-loop portion of the system acts as a vernier control to reduce any residual errors.

(3) Compensatory which implies an operation on a perceived error between the actual vehicle motion and the desired motion or input quantity. This type of control is used in the blocks labeled Quasi-linear Compensatory Control in figure 1. Compensatory differs from pursuit in that the errors only are the basis for control and command inputs are not used to structure a feedforward to the driver's output.

The levels of control activity shown are very general concepts which, nonetheless, can be applied to make predictions of system behavior. Their applicability to automobile control is limited mostly by the analyst's ability to isolate the pertinent motion error cues from the visual field geometry so that an effective controlled element can be defined.

The controlled element contains the dynamics of the vehicle and the steering system, as well as the geometry of the visual field from which the driver must extract the guidance and control cues. Driver steering control, as well as various disturbance inputs from the environment (e.g., highway gusts and roadway roughness), act on the controlled element. The resultant controlled element motions are fed back to the driver.

The roadway environment provides inputs to the system, including both commands to be followed and disturbances to be regulated against. The command input structure includes vehicles and other obstacles on the roadway, a roadway spectrum for those which curve in a random-appearing way, and discrete or deterministic roadway bends. Portions of each of these combine to give the desired path or trajectory in a given driving situation.

All the blocks in figure 1 will ultimately contain a frequency or time function. In fact, the essence of this approach is the derivation and exercise of these functions, their interaction during various driving situations, and the drawing of implications therefrom. The operational blocks (those with inputs and outputs) contain frequency functions derived from a linear or quasi-linear set of differential equations. The functions in the input blocks result from such things as spectral analyses of environmental data or time domain approximations with known deterministic forms.

Before a more specific analysis and definition of the various possible feedback loops can be made, it is necessary to have a complete understanding of the dynamic characteristics of the vehicle and the driver alone. This is accomplished below, and then attention is again focused on the closed-loop structure.

VEHICLE DYNAMICS

The dynamics of the vehicle are an important part of the effective controlled element dynamics as shown in figure 1. The driver/vehicle/roadway model examined here considers the directional or steering dynamics of the passing vehicle in three degrees of freedom. These degrees of freedom are lateral (or side) velocity v , heading angle ψ of the total vehicle, and the sprung mass (body) roll angle ϕ . Effects on the directional characteristics due to acceleration or braking in the longitudinal axes are included implicitly; i.e., the stability derivatives in the directional equations are computed for some known applied torque at the front and rear wheels. Road crown, grade, and superelevation are handled in a similar way by computing the vehicle equations for a nominal set of conditions or an operating point. This approach allows the use of linearized equations and analyses that allow greater insight into the control problems, yet provide good accuracy for vehicle side accelerations less than 0.3g (ref. 13).

The Laplace transformed directional equations in three degrees of freedom are given by the following matrix equation (1):

$$\begin{bmatrix} s - (Y_V + Y_{Vg}) & \frac{m_s e}{m} s^2 - Y_\phi & U_0 - Y_r \\ \frac{m_s e}{I_\phi} s - L_{Vg} & s^2 - L_p s - L_\phi & \frac{I_{\phi z}}{I_\phi} s - L_r \\ -(N_V + N_{Vg}) & \frac{I_{\phi z}}{I_{zz}} s^2 - N_p s - N_\phi & s - N_r \end{bmatrix} \begin{bmatrix} v \\ \phi \\ r \end{bmatrix} = \begin{bmatrix} Y_{\delta_w} \\ 0 \\ N_{\delta_w} \end{bmatrix} \delta_w + \begin{bmatrix} -Y_{Vg} \\ -L_{Vg} \\ -N_{Vg} \end{bmatrix} v_g \quad (1)$$

where v is the lateral velocity of the unsprung mass measured at the total vehicle c.g.

ϕ is the roll angle of the sprung mass about the tilted roll axis

r is the yaw rate of the unsprung mass

s is the Laplace transform variable

Y_V, Y_r, Y_{δ_w} are a result of tire side force characteristics

Y_{Vg} is a result of aerodynamic force characteristics

Y_ϕ is a combination of roll steer and camber effects

L_{Vg} is a result of aerodynamic force characteristics

L_ϕ is a combination of gravity and suspension spring effects

L_p is due to the shock absorbers and suspension friction

L_r is equal to $-m_s e U_0 / I_\phi$

N_V, N_r, N_{δ_w} are a result of tire side force characteristics

N_{Vg} is a result of aerodynamic force characteristics

N_ϕ is a combination of roll steer (landing to tire side force), camber side force, and suspension spring effects

N_p is due to shock absorbers and suspension friction

m is the total mass

m_s is the sprung mass

I_ϕ is the roll moment of inertia

I_{zz} is the yaw moment of inertia

$I_{\phi z}$ is the product of inertia coupling roll and yaw

e is the perpendicular distance from the roll axis to the sprung mass c.g.

U_0 is the forward velocity

δ_w is the mean front wheel steer angle about the kingpin

v_g is a lateral velocity gust input

The appropriate motion quantities are shown in figure 2. This matrix equation of motion provides vehicle motion quantities as perturbations of body-fixed axes. To obtain the vehicle motions in inertial space (e.g., along a roadway), it is necessary to transform the body axis motion quantities into appropriate inertial reference axes. The lateral velocity relative to an inertial coordinate system initially coincident with the unperturbed body axes is given by the integral of the lateral acceleration of the c.g., i.e.:

$$\begin{aligned} v_I(t) &= \int a_y(t) dt \\ &= \int [\dot{v}(t) + U_o r(t)] dt \\ &= v(t) + U_o \psi(t) \end{aligned} \quad (2)$$

In Laplace transform notation this becomes:

$$v_I(s) = v(s) + U_o \frac{r(s)}{s}$$

Dividing the inertial lateral velocity by U_o gives path angle γ , which is the sum of the heading angle and sideslip, i.e.:

$$\gamma = \frac{v_I}{U_o} = \beta + \psi \quad (3)$$

The lateral position relative to inertial coordinates is the integral of the inertial lateral velocity, i.e.:

$$y_I(s) = \frac{sv(s) + U_o r(s)}{s^2} \quad (4)$$

This gives the lateral position relationship between the car and the road.

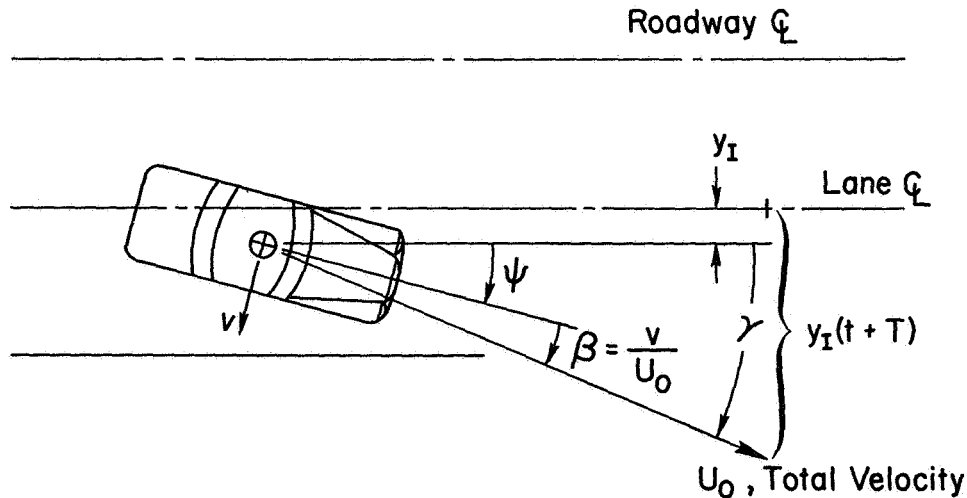


Figure 2. —Motion quantities for directional control.

Lateral acceleration at the driver's position in the vehicle, a_y , is the sum of the inertial acceleration at the axis system origin \dot{v}_I plus the effect of yawing and rolling accelerations acting through moment arms to the driver's position, i.e.:

$$a_y = \dot{v}_I + l_x \dot{r} + (l_x \sin \lambda - l_z \cos \lambda) \dot{p} \quad (5)$$

where l_x is the distance in the positive x direction from the total vehicle c.g. to the driver
 l_z is the distance in the positive z direction from the axis system origin to the driver
 λ is the sprung mass roll axis tilt angle

Laplace transforming equation (5) and introducing $v_I(s)$ gives:

$$a_y = sv(s) + (l_x \sin \lambda - l_z \cos \lambda) sp(s) + (l_x s + U_0)r(s) \quad (6)$$

It is common practice in stability and control analyses to solve the matrix equation (eq. 1) to obtain ratios of motion variables to selected input quantities. The resultant ratios of polynomials in "s" are called transfer functions. Three types of polynomials are obtained—the denominator which is always the determinant of the left-hand side of equation (1), numerators which are the determinants obtained by substituting one of the right-hand column vectors into the determinant of the left-hand side, and coupling numerators which are the determinants obtained by substituting the two right-hand column vectors into the determinant of the left-hand side in various ways. The coupling numerators are used in multiloop analyses (e.g., gust regulation) to obtain the effect on outer-loop numerators of inner-loop closures utilizing another control means. Their use is discussed in detail in reference 14 and is illustrated in reference 3.

A complete summary of the vehicle dynamics pertinent to driver control is given in reference 1. It includes the derivation of the equations of motion, definition of the vehicle transfer functions, compilation of dynamic data from a number of sources, and the calculation of numerical results for a typical American sedan. Both longitudinal and directional dynamics are included, with the emphasis on the latter. Most of the details of the vehicle dynamics are omitted here, and the needed information is simply extracted from reference 1 when required.

Vehicle transfer functions have been computed in reference 1 for a typical medium-sized American sedan weighing about 4000 pounds at speeds of 30, 60, and 75 mph. The transfer functions are represented symbolically as follows:

$$\text{Lateral velocity:} \quad \frac{v}{\delta_w}(s) = \frac{N_{\delta_w}^v(s)}{\Delta(s)} \quad (7)$$

$$\text{Heading angle:} \quad \frac{\psi}{\delta_w}(s) = \frac{1}{s} \frac{r}{\delta_w}(s) = \frac{N_{\delta_w}^r(s)}{s\Delta(s)} \quad (8)$$

$$\text{Roll angle:} \quad \frac{\phi}{\delta_w}(s) = \frac{N_{\delta_w}^\phi(s)}{\Delta(s)} \quad (9)$$

where $N(s)$ is a numerator polynomial in s and $\Delta(s)$ is a denominator polynomial in s . The numerator and denominator polynomials for the three forward speeds are given in table 1 taken from table XI of reference 1. A fairly significant difference can be seen in the dynamics between 30 and 60 mph, while there is little difference between the results at 60

and 75 mph. This indicates that one set of directional dynamics (i.e., one operating point) can be used to analyze typical high speed driving maneuvers that involve speed changes of no more than about 15 to 20 mph.

Analysis has shown that the aerodynamic forces and moments have only a negligible effect on the steer angle transfer functions of table 1. (They are of course dominant in the gust response transfer functions, through which a highway gust disturbance input acts on the vehicle—see reference 3 where the equations and transfer functions required to introduce crosswind gusts into the driver/vehicle system are derived and summarized.)

The steering system dynamics are an important part of the vehicle dynamics and the effective controlled element. They relate the driver's steering wheel movement to the steer angle of the front wheel about the kingpin. Although the theory is well understood (e.g., ref. 15), dynamic data on contemporary steering mechanisms is very sparse. Consequently, the steering system dynamics are assumed to be a pure gain for purposes of this discussion. The effects of possible steering lags on closed-loop control are considered in reference 3.

DRIVER DYNAMIC RESPONSE

All phases of driving require some form of driver control operation. As the tasks become more demanding, the driver may change his dynamic characteristics or may alter the system structure (close other loops) to obtain the required increase in control fidelity. The driver's closure of feedback loops modifies the effective dynamics of the vehicle or controlled element and in turn determines maneuver times, stability margins, and transient response characteristics. The possible feedback loops he can introduce are determined by the sensory information available.

An introductory discussion of driver response characteristics was illustrated in figure 1, and several possible types of driver response were presented: quasi-linear compensatory control, pursuit control, and internally generated maneuver commands. The first requires a fairly complex description, but it is well understood at the current time and probably comprises a significant portion of the driver's active control efforts. The remaining two are somewhat easier to describe qualitatively, but predictive models for these processes are not yet well developed.

The quasi-linear blocks in the driver/vehicle model are most appropriate for defining his response to random-appearing external inputs. This includes command inputs due to the bends and curves in the roadway or desired path, as well as disturbance inputs due to gusts and roadway roughness. These quasi-linear blocks are relatively quiescent in the presence of deterministic inputs when the other types of response dominate. They are active during pursuit control if there are also random-appearing disturbances present, which the driver cannot preview.

QUASI-LINEAR COMPENSATORY CONTROL CHARACTERISTICS.—The quasi-linear describing function model of the operator has resulted from an exhaustive series of human operator dynamic response measurements made over a period of about two decades, culminating at present in references 16 to 19, for example.

This model consists of a describing function component with parameters that depend on the system and situation, an additive remnant, and a set of adjustment rules that tell how to adjust the describing function parameters. This quasi-linear model is depicted in the illustrative single-loop block diagram of figure 3.

In its most complete form the describing function contains a gain, an indifference threshold, a time delay, an equalization characteristic, and high frequency neuromuscular

TABLE 1.—TRANSFER FUNCTIONS FOR TYPICAL AMERICAN SEDAN

Polynomial	30 mph	60 mph	75 mph
Lateral velocity numerator, $N_{\delta_W}^V(s)$	$180(s+.188)[s^2+2(.305)(6.38)s+(6.38)^2]$	$180(s+.50)[s^2+2(.284)(6.94)s+(6.94)^2]$	$180(s+.13.1)[s^2+2(2.75)(7.00)s+(7.00)^2]$
Yaw rate numerator, $N_{\delta_W}^r(s)$	$24.3(s+.49)[s^2+2(.278)(6.66)s+(6.66)^2]$	$24.3(s+.35)[s^2+2(.284)(6.95)s+(6.95)^2]$	$24.3(s+.40)[s^2+2(.279)(7.03)s+(7.03)^2]$
Roll angle numerator, $N_{\delta_W}^{\phi}(s)$	$-49.04[s^2+2(.570)(6.85)s+(6.85)^2]$	$-49.04[s^2+2(.284)(6.85)s+(6.85)^2]$	$-49.04[s^2+2(.228)(6.85)s+(6.85)^2]$
Lateral acceleration at origin numerator, $N_{\delta_W}^{a_{y_0}}(s)$	$180[s^2+2(.194)(6.73)s+(6.73)^2][s^2+2(.499)(7.43)s+(7.43)^2]$	$180[s^2+2(.293)(6.91)s+(6.91)^2][s^2+2(.160)(7.24)s+(7.24)^2]$	$180[s^2+2(.279)(6.77)s+(6.77)^2][s^2+2(.125)(7.39)s+(7.39)^2]$
Characteristic denominator, $\Delta(s)$	$.684[s^2+2(.296)(7.09)s+(7.09)^2][s^2+2(.986)(9.17)s+(9.17)^2]$	$.684[s^2+2(.785)(5.84)s+(5.84)^2][s^2+2(.290)(6.97)s+(6.97)^2]$	$.684[s^2+2(.670)(5.306)s+(5.3)^2][s^2+2(.309)(6.92)s+(6.92)^2]$

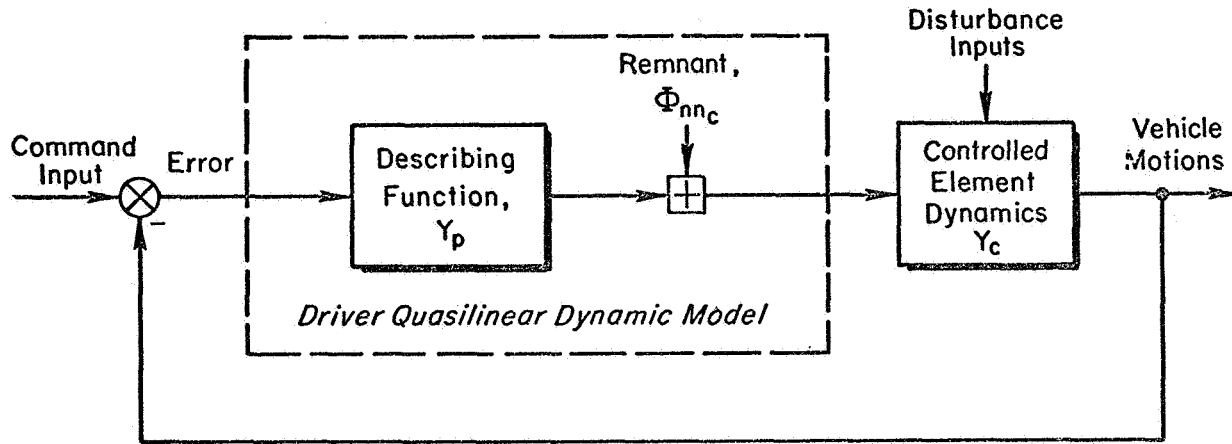


Figure 3.—Driver control in compensatory task.

system dynamics. The indifference threshold is a higher order effect that can often be ignored when the inputs are large, and under other conditions it can be accounted for by using decreased driver gain. The neuromuscular system dynamics are based on very low and very high frequency data, and can be approximated at the midfrequencies of interest in driving as a first-order lag or even as an added increment to the time delay. With these simplifications, the general driver describing function of reference 3 reduces to:

$$Y_p \doteq \frac{K_p e^{-j\omega\tau} (T_L j\omega + 1)}{(T_I j\omega + 1)(T_N j\omega + 1)} \quad (10)$$

or, alternatively,

$$Y_p \doteq K_p \left(\frac{T_L j\omega + 1}{T_I j\omega + 1} \right) e^{-j\omega(\tau + T_N)} \quad (11)$$

where K_p is the gain

$\left(\frac{T_L j\omega + 1}{T_I j\omega + 1} \right)$ is a simplified equalization characteristic

τ is the time delay

T_N is the neuromuscular system time constant

The variable $j\omega$ has been used in place of s to indicate that the describing function is most appropriate when the inputs approximate stationary random processes. The form of equation (11) is completely adequate for typical driver/vehicle closed-loop analyses involving reasonably stable vehicles. Thus, the describing function form reduces to a transfer-function-like element containing a gain, an adjustable equalization, and an exponential term

that affects only the phase angle. It is valid for a variety of drivers, inputs, vehicle dynamics, steering system characteristics, and loop structures. Most of the parameters are adjustable as needed to make the vehicle motions follow the command input and regulate against the disturbance input. Experimental measurements of all the parameters are reported in reference 3 and summarized in succeeding paragraphs.

The pure time delay represented by the $e^{-j\omega\tau}$ term is due to sensor excitation (the retina in the visual case), nerve conduction, computational lags, and other data processing activities in the central nervous system. It contains components that are closely related to certain kinds of classical reaction times. A portion of the time delay is currently taken to be a constant, because it appears to be essentially invariant with the input and vehicle dynamics for either single or dual random-appearing input tasks. However, both intersubject and intrasubject variations occur, and observed values of the constant part of τ range from about 0.05 to 0.125 second, with a nominal value of about 0.10 second (ref. 16). The remaining component of the pure time delay is an increment present when low frequency lead equalization is generated by the driver. For instance, with first-order lead equalization the base value of τ is increased by about 0.15 second.

The neuromuscular time constant T_N is partially adjustable for the task. The nature of the adjustment is input-adaptive, consisting of a monotonic decrease in T_N with increasing forcing function bandwidth ω_i . The data of reference 16 show that the incremental reduction is approximately

$$\Delta T_N \doteq 0.08\omega_i, \quad \omega_i \leq 4 \text{ rad/sec} \quad (12)$$

The observed variation of T_N with forcing function bandwidth ranges from less than 0.1 second to almost 0.5 second. In cases where the input bandwidth is not known, typical values of T_N near 0.1 second are sometimes used.

The equalizing characteristics, $(T_L j\omega + 1)/(T_I j\omega + 1)$, coupled with gain K_p are the major adaptive elements of the human that allow him to control many differing dynamic devices. Their function is the modification of the stimulus signal into a suitable neuromuscular command which is properly scaled and phased for proper overall man/machine system operation. For given input and vehicle dynamic characteristics, the form of the equalizer is adapted to compensate for the vehicle dynamics and the driver's pure time delays. The major cost of equalization is the increase in time delay incurred when low frequency lead is needed as part of the compensation.

The major adjustment rules for the equalization characteristics, evolved as generalizations of many experiments, are that a particular equalization is selected from the general form $K_p(T_L j\omega + 1)/(T_I j\omega + 1)$ so that the following properties are attained:

- (1) The driver/vehicle system can be stabilized by proper selection of gain, preferably over a very broad region.
- (2) The amplitude ratio of the product of the driver describing function and the vehicle dynamics $|Y_p Y_c|$ has approximately a -20-decibel/decade slope in the crossover region—that frequency band centered on the crossover frequency ω_c .
- (3) $|Y_p Y_c| \gg 1$ at low frequencies to provide good low frequency closed-loop response to system commands and good suppression of the effects of disturbances.

The region near the crossover frequency ω_c (where $|Y_p Y_c| \doteq 1$) has a surpassing importance in the synthesis of feedback systems. First, for good performance the crossover frequency must exceed the largest frequency ω_i at which there are appreciable components of the commands or external disturbances. In this sense, the crossover frequency amounts to an upper bound on the frequency region wherein $|Y_p Y_c| \gg 1$, a condition that

provides good following of the system input by the output and suppression of disturbance effects. Second, the shape of $Y_p Y_c$ at and near the crossover frequency determines the dynamics of the dominant modes of the system response. The conditions for neutral oscillatory stability are

$$|Y_p Y_c| = 1 \text{ and } \angle Y_p Y_c = -\pi \text{ (in radians)} \quad (13)$$

when sign $Y_p Y_c(0)$ is plus. Thus, neutrally stable or unstable dominant modes are usually avoided if the system satisfies the condition

and

$$|Y_p Y_c| < 1 \text{ when } \angle Y_p Y_c = -\pi$$

$$\angle Y_p Y_c > -\pi \text{ when } |Y_p Y_c| = 1$$

These are the often quoted criteria of positive gain margin and phase margin. For simple minimum phase or minimum phase plus transport lag systems, they are an expression of the Nyquist stability criterion. Because stability is a fundamental requirement in any feedback control system, these requirements for gain and phase margin may severely restrict the choice of the crossover frequency.

The human corresponds fairly well to the criteria for a good controller, varying somewhat with vehicle dynamics and input bandwidth. Experimental values of crossover frequencies ω_c and phase margins ϕ_M for several dynamic forms and bandwidths are given in table 2, abstracted from reference 16. These data indicate the invariance properties of ω_c , which are:

$\omega_c - K_c$ independence.—After initial adjustment, changes in vehicle gain K_c are offset by changes in driver gain K_p so that system crossover frequency ω_c is invariant with vehicle gain.

TABLE 2.—PHASE MARGINS AND CROSSOVER FREQUENCIES

Y_c	$\omega_1 = 1.5 \text{ rad/sec}$		$\omega_1 = 2.5 \text{ rad/sec}$		$\omega_1 = 4.0 \text{ rad/sec}$	
	ϕ_M (rad)	ω_c (rad/sec)	ϕ_M (rad)	ω_c (rad/sec)	ϕ_M (rad)	ω_c (rad/sec)
K_c	0.51	5.1	0.75	5.7	0.75	6.7
K_c/s	0.42	4.6	0.73	4.7	0.94	5.0
$K_c/(s-2)$	0.35	4.6	0.66	5.0	0.70	5.2
K_c/s^2	0.26	3.2	0.51	3.3	0.73	1.8

ω_c — ω_i independence.—System crossover frequency depends only slightly on input bandwidth for $\omega_i < 0.8 \omega_{c0}$, where ω_{c0} is the ω_c for a given Y_c extrapolated to zero input bandwidth ($\omega_i = 0$).

ω_c regression.—When ω_i nears or becomes greater than $0.8 \omega_{c0}$, the crossover frequency regresses to values much lower than ω_{c0} .

Phase margin adjustment with ω_i .—Because ω_c is essentially independent of ω_i and ΔT_N is directly proportional to ω_i , the system phase margin ϕ_M is directly proportional to ω_i . This strong dependence of phase margin on the input bandwidth is associated with the linear variation of ΔT_N with ω_i and is essentially an alternate statement of equation (12).

The crossover frequencies presented in table 2 are maximum values for skilled subjects in fixed-base simulators without motion feedbacks. These will ordinarily be considerably reduced during driving, especially when full attention is not demanded. The maximum extent of this reduction can be readily estimated for conditionally stable systems on the basis of stability alone. To obtain an estimate of the full-attention driving value for vehicles that are always stable, consideration should be given to:

- (1) Gain reduction
- (2) τ_e increase, due to conflicting demands
- (3) τ_e increase, due to steering dynamics, especially when holding trim loads against road crown or steady crosswinds
- (4) Statistical variation in ω_c with time and between subjects

This leads, in practice, to larger phase and gain margin criteria, typical examples of which are used in the driver/vehicle loop closures shown below and in reference 3.

In many driving situations the driver's regulation or control activity is only intermittent, so the average crossover frequencies will invariably be less than those estimated for full-attention driving. This reduction is probably due primarily to an increase in indifference threshold, and thus may not result in much of a reduction for large amplitude motions. A matter of some importance is the spread between the highest and lowest ω_c values possible (the lowest often merges with the unattended condition), which is a measure of the degree of vehicle configuration forgiveness. This or an associated measure also implies limitations on indifference threshold, sampling, minimum and maximum average movements per second, minimum and maximum information rates, etc.

Simplified Crossover Model for the Driver Describing Function.—Many of the preceding remarks about the rationale of equalization adopted by the driver can be simplified by using the approximate crossover model. The experimental data and consideration of the requirements of good feedback system performance both lead directly to the conclusion that the driver adjusts his describing function so that the open-loop function $Y_p Y_c$ in the vicinity of the gain crossover frequency ω_c is closely approximated by

$$Y_p Y_c \doteq \frac{\omega_c e^{-j\omega \tau_e}}{j\omega} \quad (14)$$

where τ_e is an effective pure time delay that includes the neuromuscular time constant T_N as well as τ and any net high frequency controlled element lag. This crossover model is not a replacement for the more complete model, but it is a convenient engineering approximation. While it is a better description of amplitude ratio characteristics than of phase characteristics, it often describes the most significant features of operator behavior adequately. This is because the actual shape of the open-loop function away from the gain crossover frequency is usually almost irrelevant to the closed-loop performance. It is, therefore, often unnecessary to retain terms in the describing function whose influence is not felt in the immediate vicinity of gain crossover.

Table 3 shows the application of the adjustment rules to the prediction of a driver's describing function and to the formulation of the crossover model for several simple (limiting-case) controlled element dynamics. In many instances more complicated controlled elements can be satisfactorily approximated in the vicinity of gain crossover by these simple limiting cases.

Driver Remnant Characteristics.—The remnant is that portion of the driver's output which is not linearly correlated with the input. It is always present to some extent, and it is probably the most significant undesired input into the steering system from a practical standpoint. The remnant is considered to be a random process added to the output of the driver's describing function block to form the total driver's output. It is denoted by a power spectral density Φ_{nn} in figure 3. The point of application could be moved to other places in the loop as long as no nonlinear elements are passed in the process. The remnant can have significant power at frequencies that are high enough to excite lightly damped high frequency steering system modes present under conditions of low Coulomb friction (e.g., ref. 15).

The major source of remnant appears to be nonstationarity in the operator's behavior (ref. 16), manifest as time-varying components in gain K_p and effective time delay τ_e . The remnant component due to τ_e variation is usually increased when low frequency lead is generated by the driver. Also, low frequency lead is associated with a pulsing behavior by the operator in which his output tends to be pulses with areas roughly proportional to the stimulus amplitude. This is an additional remnant source for controlled element dynamics that require low frequency lead equalization. An extensive body of remnant data shows that the remnant power increases with the order of the controlled element dynamics and with increases or decreases in controlled element gain away from an optimum value. It decreases with input bandwidth.

For nominally good dynamics (e.g., $Y_c \doteq K_c/s$) the remnant power is about 30 decibels down relative to the input. Consequently, the remnant can usually be neglected from the standpoint of predicting driver/vehicle closed-loop response characteristics when the vehicle characteristics are reasonably good. This assumption is made in conventional analyses as a starting point. Then, if there is reason to believe that significant remnant power is present in the vehicle output motion quantities, the remnant can be added to the output of the driver/vehicle system.

Driver Response in Multiple-Loop Situations.—The closed-loop structure of figure 1 involving angle and path feedbacks is called multiloop single-point control. Multiloop implies two or more dynamically coupled motion variables, and single-point refers to one control means, the steering wheel. Experimental measurements of operator response in such a multiloop situation (ref. 19) showed that his describing function in the outer loop can be obtained by application of the single-loop model described above. Thus, the equalization in an outer loop involving lateral position is obtained by direct application of the describing function model once the appropriate inner loop has been closed.

Ordinarily the inner feedback loops supplied by the driver act as parallel equalization for the outer loop, or provide feedbacks or crossfeeds that suppress subsidiary controlled element degrees of freedom that have undesirable effects on subsequent loops. Because the role of the inner loops is so dependent on outer-loop requirements, the rules cited above for the single-loop model are not generally applicable, e.g., even stability of an inner loop may not be required. The types of inner loops closed and the equalization selected should be compatible with one or all of the following considerations:

(1) Outer-loop adjustments per the single-loop adjustment rules become more feasible, e.g., $|Y_p Y_c|$ for the outer loop can be made approximately -20 decibels/decade with less outer-loop equalization by the driver.

TABLE 3. — DRIVER DESCRIBING FUNCTION AND Crossover MODEL

WHEN THE EFFECTIVE CONTROLLED ELEMENT TRANSFER FUNCTION IS (Y_C)	THE DRIVER'S APPROXIMATE DESCRIBING FUNCTION IS (Y_P)	THE Crossover MODEL IS ($Y_P Y_C$)	EFFECTIVE TIME DELAY FOR $\omega_1 = 0$ $[\tau_e(0)]$ (sec)
$\frac{K_C}{s} = \frac{K_C}{j\omega}$	$K_P e^{-j\omega\tau_e}$	$\frac{K_P K_C e^{-j\omega\tau_e}}{j\omega}$	$1/3$
K_C	$\frac{K_P e^{-j\omega\tau_e}}{(T_I j\omega + 1)} ; \frac{1}{T_I} \ll \omega_C$	$\frac{K_P K_C}{T_I} \frac{e^{-j\omega\tau_e}}{j\omega}$	$1/3$
$\frac{K_C}{s^2} = \frac{K_C}{(j\omega)^2}$	$K_P e^{-j\omega\tau_e} (T_L j\omega + 1) ; \frac{1}{T_L} \ll \omega_C$	$\frac{K_P K_C T_L e^{-j\omega\tau_e}}{j\omega}$	$1/2$

(2) The sensitivity of the closed-loop characteristics to changes in either inner- or outer-loop driver characteristics is reduced from that in an outer-loop-only situation. This includes the improvement of stability margins.

(3) The loop structure and equalization selected are those for which total subjective opinion rating is the best obtainable.

It frequently happens that the driver describing function synthesized for the inner loop alone via the single-loop model is also the one that best enhances the outer-loop closures. This is most common in situations where the basic vehicle possesses good stable dynamic characteristics, and the control task is merely following command inputs or suppressing disturbances — not stabilizing the vehicle.

PURSUIT CONTROL CHARACTERISTICS. —The pursuit control block in figure 1 operates on the input using the driver's preview of a desired path, rather than a perceived motion or path error as in the compensatory case. Recent experiments (ref. 13) have shown that the describing function magnitude of the driver's pursuit feedforward block Y_{p_i} is approximately equal to the inverse of the magnitude of the effective controlled element dynamics, i.e., $|Y_{p_i}| = |1/Y_c|$. The net effect is to make the amplitude of the total closed-loop describing function from the command path to actual vehicle path be approximately unity (although the data show some variations in phase angle), and the commanded path and the actual path are approximately equal. The driver must be able to see and use the input itself to structure this block. The compensatory loops can be active when the pursuit loop is operative, providing, for example, vernier corrections and regulation against disturbance inputs that are most evident in the error.

INTERNALLY GENERATED MANEUVERS. —When a discrete disturbance (e.g., an isolated gust) is encountered, the change in the vehicle motion quantities is perceived by the driver as a motion error. The well-trained driver will recognize that the disturbance has a deterministic form for which there is some appropriate skilled response. If the input is a step, for example, he may make a classic response (e.g., ref. 21) involving a time delay, a relatively rapid steering wheel motion during a rise time phase, and an error-correcting phase. Recent operator response modeling activities (ref. 22) have concentrated on deriving and refining models for discrete inputs (such as steps and ramps) with various effective controlled element dynamics. Some potentially useful results have been obtained, but their application to the prediction of driver behavior is not yet state of the art. Any discrete response block is assumed to be quiescent in the presence of low frequency random motions, and to respond only to simple deterministic signals. During such a response the other driver blocks are assumed to be relatively inactive.

A number of the control actions made by a practiced driver during various phases of driving are done in a precognitive (or open-loop) manner. A good example is pulling out to pass and pulling back in. In this case the driver produces the appropriate steering action based on an internally generated pattern previously evolved during a learning process. He is cognizant of the effective controlled element, the command input, and the state of the motion quantities reflected in the error; but is not operating on any motion quantity in a linear or quasi-linear way. Safety aside, he could in effect close his eyes and complete the maneuver, leaving some residual position error at the end. The patterns produced may result in nearly optimal response according to some (yet undefined) criterion such as minimum time, minimum overshoot, etc. The essence of this block for maneuver command generation is that it produces whatever control signal is required to accomplish the desired maneuver to some degree of accuracy acceptable to the driver.

STRUCTURES OF DRIVER/VEHICLE CLOSED-LOOP SYSTEMS

Perhaps the most difficult problem in driver/vehicle closed-loop analysis is to determine what sensory feedback loops the driver is using. There are two general approaches to deriving the closed-loop structure: from consideration of perception and from guidance and control theory. The perceptual basis relies on in situ experimental observation of subjects driving, and can require extensive instrumentation, eye-movement cameras, highly structured and constrained experimental environments, etc. The primary aim is to discover experimentally what cues are used to drive. The control theory approach is to consider the driver/vehicle system as a guidance and control problem, and then to develop the kinds of vehicle motions that must be sensed and commands that must be inserted to satisfy the guidance and control needs. The guidance and control problem does not have a unique solution, so the initial results of this kind of study are a number of sufficient systems. Consideration must then be given to the driver's ability to perceive (sense) the vehicle motion and input quantities and to the nature of the operations required on the sensed quantities in closing the driver/vehicle system loops. Those potential systems that involve readily sensed quantities and relatively simple (e.g., proportional, minimum conscious effort, etc.) operations on the sensed motion variables are then accepted as good candidates. The control theory approach is relatively straightforward, given the current state of knowledge of typical operator/vehicle control laws and closed-loop analysis techniques. The best approach is to marry the guidance and control and the perceptual theories, for the net results must be of necessity compatible, and the two viewpoints offer much to one another as collaborators and corroborators. However, it has been necessary to rely on the control theory approach as the main tool here because of the limited amount of perceptual data relevant to vehicular control.

SINGLE-LOOP STRUCTURE. —The problem of how the driver perceives the feedback variable is secondary in the deductive guidance and control approach. The objective is to determine how well the variable permits the driver to control the vehicle if it can be sensed. This is pursued by attempting to discover likely feedback loops that may exhibit good characteristics as either a command loop or as a subsidiary loop that provides equalization for a command loop. It is also desired to identify those loops which possess poor characteristics as single-loop systems that either rule them out as possible feedbacks or indicate the need for inner-loop equalization or that exhibit properties that make them candidates as accident causes if they are inadvertently opened or closed.

A systematic search for likely feedback loops has been accomplished (ref. 3) using the dynamics from reference 1 of a typical sedan (circa 1965) and the driver describing function characteristics. The search took the form of a survey of single-loop closures directed toward determining those loops which the driver might close. The quality of the resultant closures was judged by

- (1) The equalization required by the driver to provide a stable system consonant with the adjustment rules

- (2) The closed-loop performance attainable as measured by crossover frequency, stability margins, etc.

- (3) The subjective opinion the driver would have of his control task (as determined by the amount of lead equalization he must adopt, precision of describing function adjustment required, etc.)

The closures were restricted to compensatory mechanizations. This is valid and consistent with such concepts as preview when one considers that the effective controlled

element dynamics include the geometry and kinematics of the visual field as well as the dynamics of the vehicle. In essence, a guidance or control cue is perceived someplace in the visual field by the driver, and to the extent that the driver steers the vehicle to modify (e.g., follow or reduce) this cue in some way, he is acting in a compensatory manner.

The results of driver/vehicle single-loop surveys are summarized in table 4. The same closure criterion was used for each of the loops in table 4 to maximize the validity of relative comparisons. Also, the same effective time delays were used in all systems for computational convenience—this has the effect of making the crossover frequency for those systems with driver-generated low frequency lead larger than it would be with a more precise estimate. The absolute levels shown will vary from one driver to another, depending on skill, attention, fatigue, etc.; but these levels should all change in roughly the same way as a function of these factors for a given driver.

Closure of an inertial lateral deviation loop is very likely necessary to stay in the lane or to follow a desired trajectory. The driver is assumed to be steering according to his lateral deviation with respect to an inertial axis reference such as the roadway or lane centerline. Although the driver can easily perceive this motion quantity, this single-loop control mode is relatively difficult. The effective controlled element transfer function is given by:

$$\frac{y_I}{\delta_w} = \frac{180[s^2 + 2(0.160)(7.24)s + (7.24)^2][s^2 + 2(0.293)(6.91)s + (6.91)^2]}{0.684s^2[s^2 + 2(0.785)(5.84)s + (5.84)^2][s^2 + 2(0.290)(6.97)s + (6.97)^2]} \quad (15)$$

The system survey plot is shown in figure 4. The system is unstable with a pure gain plus time delay (unequalized) driver describing function, as shown by the solid lines, because of the two poles at the origin (see root locus). A relatively large amount of driver lead equalization is required to attain even minimal stability margins. A skilled driver could add low frequency lead at about 0.5 rad/sec and obtain a crossover frequency of about 1.0 rad/sec with 30° of phase margin. This is shown by the dashed lines in figure 4, and corresponds to a driver model of the form

$$Y_{py_I} \doteq \frac{K_{py_I}}{0.5}(s+0.5)e^{-0.3s} \quad (16)$$

Although a skilled driver could achieve the lead equalization of equation (16) for a short period of time, it would be difficult and he would have a poor subjective opinion of the vehicle's dynamics (ref. 23). Thus, an inertial lateral deviation loop has poor characteristics as a single-loop system, and plainly requires additional (inner) loop closures to alleviate the need for large driver lead to obtain a satisfactory driver/vehicle system bandwidth.

Although $y_I \rightarrow \delta_w$ is not a likely primary control loop because of the low bandwidth and high driver skill required, it may be important after perceptual transitions. This could occur when the driver suddenly loses his view down the road and is forced to steer on the basis of his lateral position in the lane (e.g., entering a tunnel or fog bank). Such a regression would cause him to slow down so that the lower driver/vehicle system bandwidth would still enable him to follow any likely command input.

Both heading angle and path angle are control feedbacks that offer good closed-loop characteristics. (Detailed surveys are shown in figs. 8 and 10, respectively.) Path angle

TABLE 4. - SUMMARY OF SINGLE-LOOP SURVEYS

FEEDBACK	DRIVER EQUALIZATION	ESTIMATED SYSTEM CROSSOVER FREQUENCY (rad/sec)	ESTIMATED RELATIVE DRIVER OPINION
Inertial lateral deviation $y_I \rightarrow \delta_w$	Large lead	1	Poor, because of large amount of lead equalization and low bandwidth
Heading angle $\psi \rightarrow \delta_w$	Gain only	1.5	Good
Path angle $\gamma \rightarrow \delta_w$	Gain only	1	Good, particularly with small lead equalization (ref. 3)
Lateral velocity $v \rightarrow \delta_w$	Lag-lead	1	Fair, because of low bandwidth and lag equalization
Heading rate $r \rightarrow \delta_w$	Lag-lead	2.5	Good
Lateral acceleration at driver's head $a_y \rightarrow \delta_w$	Lag-lead	1	Fair, because of low bandwidth, lag equalization, and high sensitivity

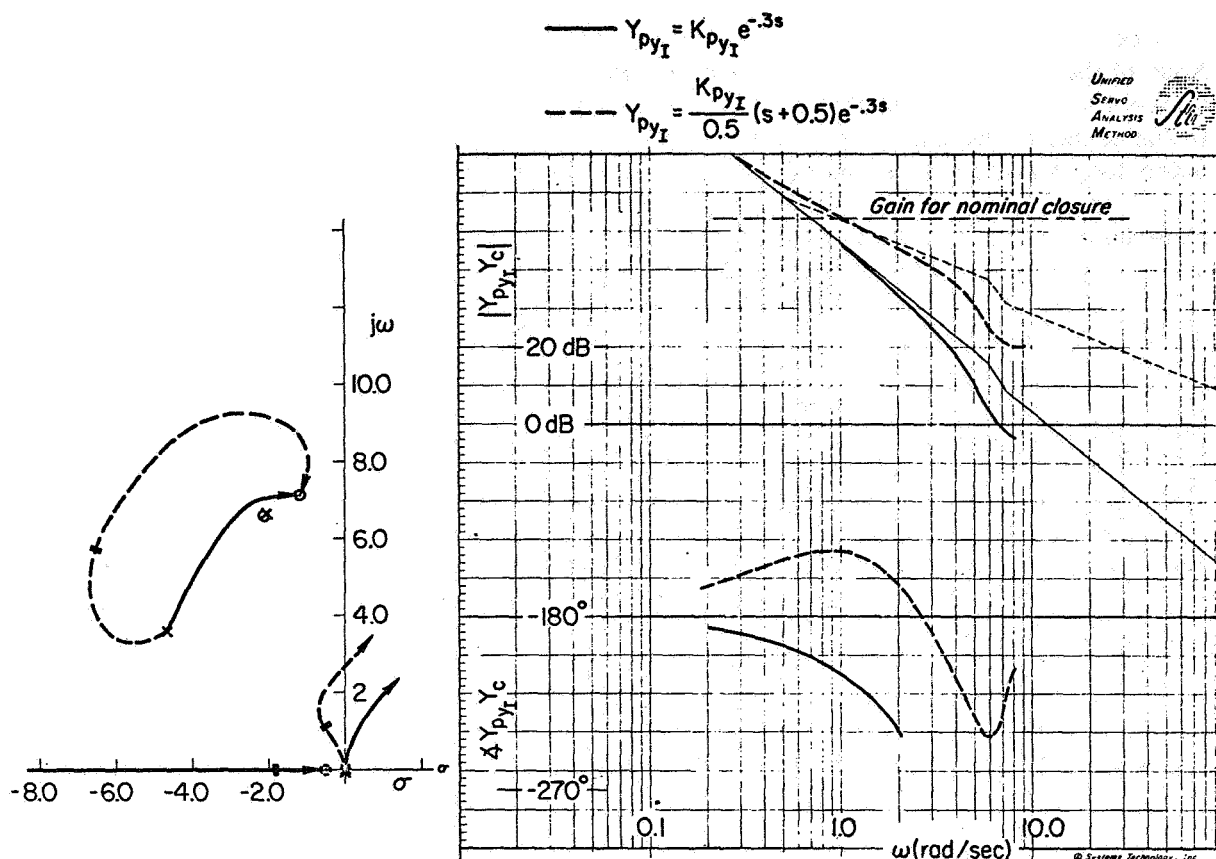


Figure 4. —Inertial lateral deviation loop closure.

control combines heading control (a good loop by itself) with body axis lateral velocity (sideslip), which is only fair taken alone. Both heading angle and path angle systems can serve as outer loops if an intermittent trim loop is employed to occasionally reduce or reset the lateral deviations. Also (as described later), either of these angle systems can serve as an inner loop to reduce the lead requirements for an inertial lateral deviation outer-loop system.

The heading rate system is also good and is a conceivably useful inner loop. Path angle rate (not shown) has similar potential. On the other hand the lateral acceleration at the driver's head (a quantity related to path angle rate) is not a particularly good system because it is highly sensitive to driver gain variations; that is, the difference between the minimum gain required to provide some control and the maximum gain permissible without instability is not large. This also places an additional requirement on the need for relatively high driver skill.

The single-loop closures in table 4 that have poor characteristics are unlikely loop closures under favorable conditions because of the poor system performance or excessive driver demands. They can, however, be important in situations where a better loop structure has been destroyed due to changes in the cues available or other disturbing factors. Under such perceptual transitions, these poor single-loop systems may momentarily exist as a transitional phase.

MULTILOOP STRUCTURES.—The good single-loop closures shown in table 4 will all provide systems with good performance in following a command input of that motion variable. None of the single-loop closure systems shown will do a very good job of following a path or trajectory command input that involves minimizing lateral position (deviation) errors to stay in the center of the lane or roadway. Thus, the need exists for the driver to augment his outer-loop structure with inner loops that serve as appropriate equalization. Among these multiloop systems, those which require little or no driver equalization (i.e., only gain plus time delay in each of the loops) are to be preferred from both a closed-loop performance and a driver subjective opinion standpoint.

Three different multiloop systems have evolved to date. Their block diagrams are shown in figure 5, and their characteristics are summarized in table 5, based on the detailed system survey analyses of the appendix. Each of the selected multiloop configurations is discussed briefly below in their nominal good configuration. Ways in which control difficulties can arise via transitions in the multiloop structures are discussed in reference 3.

The time-advanced lateral deviation structure of figure 5a assumes that the driver operates on an estimated or projected lateral deviation error. Preview here is explicitly required, because this error gives the lateral position the vehicle would have at some point ahead of the vehicle if it continued along its current path (fig. 6). The range to the point of regard is given by R , which is approximately equal to velocity U_0 times time T to travel to that point. Time advance T provides a perceptual preview that results in a pure lead equalization term in the effective controlled element dynamics. This, in turn, offsets the undesirable K_C/s^2 form of the lateral deviation dynamics at low frequency. When operating as described here, this system is essentially single-loop, although the multiloop aspect of figure 5a is useful for later studies of transitions and to make figure 5a comparable with figure 5b.

The amount of lead is given by R/U_0 , which is under the direct control of the driver and depends simply on how far down the road he is looking. The analyses in the appendix indicate that a time of 5 to 10 seconds is adequate. Larger values are of little help under normal circumstances. Values of T less than 5 seconds (440 feet at 60 mph) are not as good because they do not compensate sufficiently for the inherent lags in the driver/vehicle system. This system structure provides a relatively high crossover frequency and good lateral position control on straight roads. Some conceptual and analytical difficulties arise when it is necessary to follow a curving roadway or a passing trajectory, e.g., a guidance law or scheme is needed to provide an appropriately advanced command input to compare with the projected lateral deviation.

The path angle plus lateral deviation system of figure 5b assumes that the driver operates on these motion quantities as separate entities. It differs from the figure 5a model by having the present lateral deviation available to the driver for comparison with a desired command input and derivation of a position error. With equivalent gains the two systems are indistinguishable in the absence of a command input. Table 5 shows that the path angle plus lateral deviation system gives adequate stability and reasonably good command following. The stability of the system is relatively insensitive to joint variations in the loop gains, although the crossover frequency and (hence) system bandwidth will change of course. Consequently, the system is reasonably forgiving of momentary lapses and does not require continuous control.

The heading angle plus lateral deviation structure of figure 5c assumes that the driver operates on these motion quantities separately and combines them to produce a steer angle response. It differs only slightly from the preceding model in its use of heading angle instead of path angle (which is heading plus sideslip) as the inner loop. Table 5 shows that the heading angle plus lateral deviation system gives good command following and performance, is

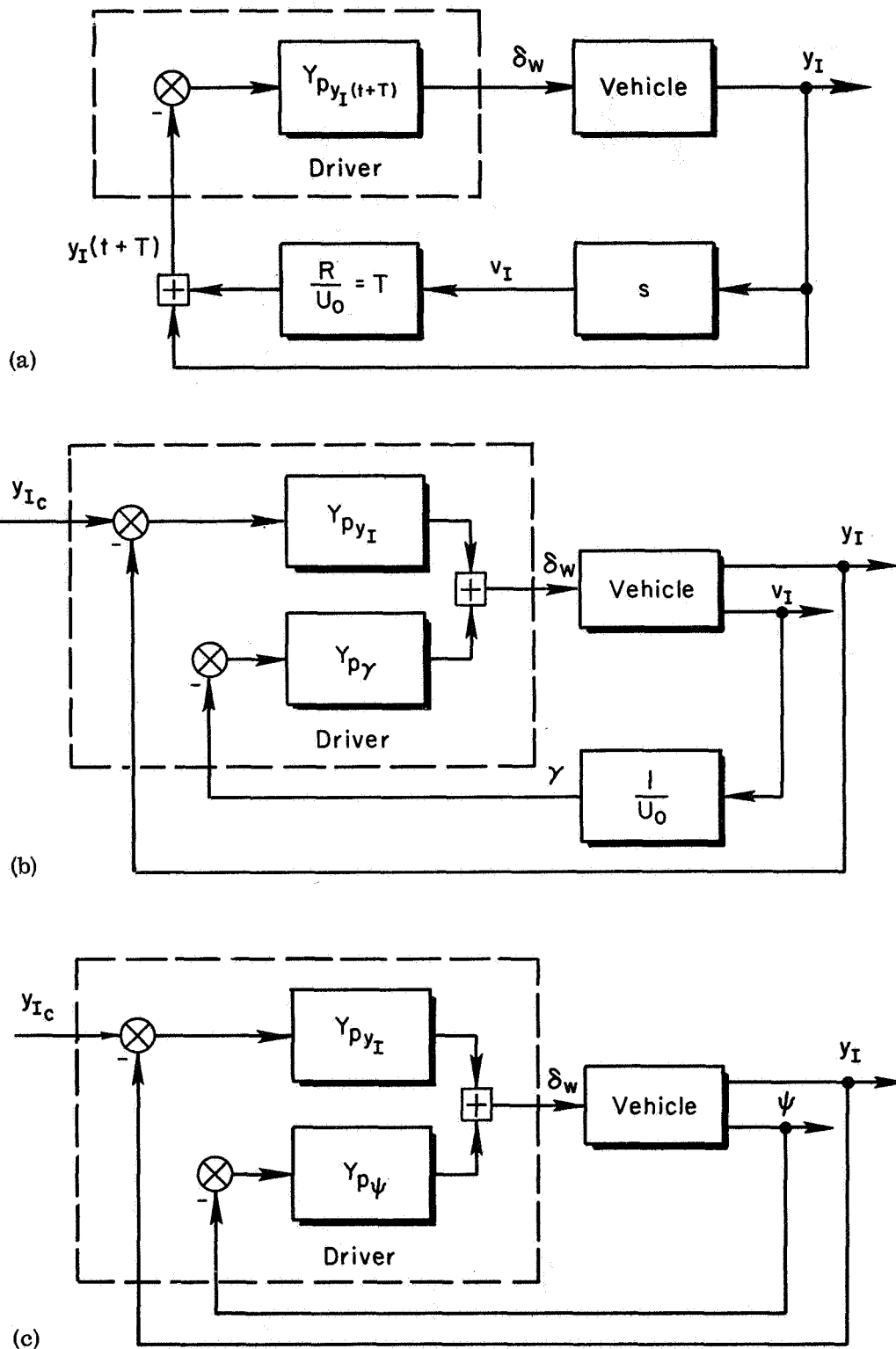


Figure 5.—Block diagrams for selected multiloop closures. (a) Inertial lateral deviation advanced in time. (b) Path angle plus inertial lateral deviation. (c) Heading angle plus inertial lateral deviation.

TABLE 5. — MULTILoop SYSTEM CHARACTERISTICS

SYSTEM	INNER LOOP		OUTER LOOP		REMARKS
	Equal- ization	Cross- over Freq. (rad/sec)	Equal- ization	Cross- over Freq. (rad/sec)	
Inertial lateral deviation advanced in time	-	-	Gain only	1.2	Fine for straight roads. Difficult to define command input form and point of entry into the system. Sensitive to changes in preview.
Path angle plus inertial lateral deviation	Gain only	1.0	Gain only	0.53	Control not sensitive to changes in driver adaptation or driver atten- tion. Outer loop can operate intermittently.
Heading angle plus inertial lateral deviation	Gain only	1.5	Gain only	0.6	Control not sensitive to changes in driver adaptation or driver atten- tion. Outer loop can operate intermittently.

relatively insensitive to changes in driver adaptation, permits fairly inattentive control, and has the attribute needed for good driver opinion. It has the additional possible advantage that heading angle (which is simply the subtended angle between reference lines on the car and in the roadway) may be easier to perceive under certain circumstances than path angle (which is the angle between a roadway reference line and the line to the point of no relative motion in the surround). Note, finally, that sideslip angles are usually relatively small, in which case heading angle and path angle are almost the same quantity.

To avoid potential confusion about the relative merits of the three systems insofar as outer-loop crossover frequency is concerned, it should be emphasized that the larger value for the advanced-in-time system is an artifact of the loop closure criteria used. As already noted, the closed-loop dynamics of the three systems can be made very similar if the loop gain and preview times are appropriately adjusted. However, in making the estimates for the systems of figures 5b and 5c, the inner loops are closed with relatively large stability margins, so as to be representative of conditions with possibly intermittent closures of the outer loop. The advanced-in-time system of figure 5a, on the other hand, has no separable inner loop, so this consideration did not apply.

None of the systems shown in figure 5 include the two feedforward channels or the discrete response feedback loop of figure 1. This is because the closed-loop systems shown in figure 5 are most appropriate for command following or regulation tasks of a reasonably continuous nature with inputs that are more or less random appearing. The outer types of response involving learned maneuvers, etc., are important and do dominate the driver's control activity in some phases of driving. However, the closed-loop type of control is fundamental to many other phases of driving and plays a key take-over role in circumstances where the learned maneuvers and patterned responses either cannot be structured or are suddenly destroyed for one reason or another. The compensatory loops are also used in early phases of learning (the unskilled driver) and under conditions of extreme stress (the startled or confused driver). If the closed-loop systems will not or cannot work, then there is little chance that the driver/vehicle system will function safely for long. Thus, understanding their form and operation gives the point of departure for either adding the other channels or switching to them as needed.

These three systems do not exhaust the possible multiloop structures that can be concocted. Yet they are the only ones found to date that satisfy the guidance and control requirements for command following and disturbance regulation with good performance, insensitivity to variations in the driver's dynamic adaptation, good predicted subjective opinion from the driver, etc. Further, they are not inconsistent with available perceptual data obtained from driving experiments on the highway. In addition to modeling the driver/vehicle system in a useful way, they can also provide a new framework for devising further experiments to study driver perception and control processes. All three structures give roughly the same performance under good conditions, although they differ in some details. The principal distinction between them lies in the way in which they degrade when the nominally good structure is disturbed. Such a disturbance might disrupt the perceptual interaction between the driver and the surroundings. It might also modify the driver element by distraction, inattention, or other degradation, the vehicle element in a mechanical way, or the roadway environment's texture, illumination, etc. The way in which these transient modifications in the dynamics of elements within the guidance and control loop(s) or in the loop structure itself can lead to dangerous vehicle motions is examined in reference 3.

CORRELATION WITH EXPERIMENTAL OBSERVATIONS.—The identification of preferred driver/vehicle loop closures and sensory cues from a control theory approach

provides a new framework for reviewing past perceptual experiments. The perceptual basis is somewhat fragmentary, but all the available evidence tends to corroborate the results of analytical deduction previously cited. The following verbal evidence and remarks from representative sources will at least illustrate the trend of activities in this area.

A pioneering series of experiments by Gordon (ref. 25) attempted to define the driver's visual input experimentally. The apparatus was a helmet-mounted aperture for one eye comprised of a tube 3.5 inches long with a variable diameter of 1 inch or less. The resultant aperture angles were 4° and 9.75° . The other eye was masked. A camera was mounted on the helmet coaxially with the aperture tube. The task was to drive along a narrow, winding two-lane country road. He found that most records showed continuous visual shifts forward to the limit of the visible road and then backward toward the vehicle. This bimodality of fixation positions suggests that the driver may be looking down the road to obtain heading or path angle information, and then occasionally looking near the vehicle to sample lateral position or deviation. His data also showed that about 60 percent of the time the driver was looking farther than 150 feet ahead of the vehicle, while about 80 percent of the time he was looking more than 100 feet ahead. At the test speeds of about 15 mph (20 fps) these are preview times of 7.5 and 5 seconds, respectively. These times correlate well with desirable preview time T found for the time-advanced lateral deviation model of figure 5a. The use of small aperture viewing was reported to cause stress, suggesting that denial of the peripheral cues may lead to a significant degradation in the perceptive structure.

In a more recent paper, Gordon (ref. 26) expands on the streamer theory (after Calvert, ref. 27), which states in essence that the driver perceives motion from objects in the visual field streaming across his field of view and emanating from a central focus (when following a nominally straight path). An alternative is taken by Gibson (ref. 28) who asserts that the focus of expansion provides the directional cue rather than the flow characteristics of the velocity field emanating from the focus. Regardless of how the driver senses the direction to this focus, this cue is precisely the path angle cue discussed previously and used in the loop closures of the appendix. Gordon goes on to say that all parts of the visual field (road borders and lane markers) move when the wheel is turned, and no one part is essential for tracking. His studies showed that the driver may assume a somewhat unlocalized surveillance of the road, which would facilitate seeing a steady state (flow field) and also reduce nystagmus. He later suggests that "the driver may become aware of the car [motion] by slewing shifts in direction, and by sideslipping sidewise movements," and "sidewise movement of the road borders is perceptible either as a movement or change in position." He concludes that "on the basis of human perception theory [alone] it is difficult to determine which of the four combinations of slew, sideslip, rate, and amplitude the driver perceives. The driver responds to a total situation, not to isolated or ranked cues."

Schmidt (ref. 29) discusses perception in general, and then notes that "the driver recognizes the movement of the car from the apparent flow or streaming of the objects in the visual field. . .," agreeing in essence with Gordon and Calvert.

Biggs (ref. 30) conjectures that in driving the central vision is occupied with the detection of obstacles in the immediate path, while the peripheral vision is employed in the task of tracking the guideline or dividing line. He notes that "the guideline is seen near the vehicle, and its lateral motion provides the dominant directional cue." It could be inferred from this that the directional cue detected in the periphery from motions of the guidelines (or streamers) yields heading or path angle and possibly angular rate. Inertial guidance is, of course, just the current position of the guideline or streamer.

Even though these references are all fairly recent, none of them specifies what loops to close. They do support, however, the theory of driver control derived herein from guidance and control considerations. While there are apparent differences between these

experimental results, some semblance of order is derived by viewing them within a guidance and control structure, and the operational entities extracted from the perceptual cues do provide the prescribed feedbacks.

SOME IMPLICATIONS OF THE APPROACH

The objectives of the research of which this paper forms a part have been to gain further understanding of driver control processes, and to determine why guidance and control errors occur and how they can be avoided to reduce the hazard. The major results of the first objective have been presented here. These have involved the derivation and development of operational models for the driver/vehicle/roadway system which use the techniques and principles of feedback control theory. The resultant models provide powerful analytical tools that can be used to:

- (1) Determine the role and relative importance of the various elements (driver, vehicle, etc.) in the operation of the overall system
- (2) Define the levels and nature of the interaction between system elements
- (3) Assess the effect on the characteristics of the overall system of changing parameters within and between the elements

In essence, a new capability has been established which can guide the highway systems engineer in his efforts to improve the driver/vehicle/roadway system, or to assess the effect on vehicle control of proposed system changes that may be thrust upon him. Although not a panacea, the approach provides fresh insights, some alternative views to the traditional, a basis for interpretation of experimental results, and a useful foundation for designing new key driver control experiments that can more fully validate the derived theory.

The second and third objectives are treated in reference 3 using example applications of the driver/vehicle closed-loop models. There the following problem areas are considered:

(1) Transitions in the driver/vehicle/roadway system, with particular emphasis on sudden changes in the perceptual structure and their varying degrees of resultant hazard. One example treated loss of driver preview in the time-advanced lateral deviation structure in the presence of a random-appearing disturbance input. Another involved failures of the path angle inner loop (while retaining the lateral deviation loop) at various points during a passing maneuver. Another example considered loss of the heading angle inner loop followed by introduction of suitable driver lead equalization in the lateral deviation outer loop.

(2) The effect of acceleration feedbacks, with emphasis on ways they can interfere with good visual loop structures. The influence of driver location (relative to the center of the vehicle) on the acceleration cue was examined first. Then transitions from visual control to steering control based on accelerations at the usual position of the driver's head were studied. Finally, inadvertent steering inputs due to arm/hand inertia of the steering wheel (in the presence of vehicle accelerations) were examined.

(3) Highway gust disturbance inputs, with attention to the types of highway gusts that can be most troublesome and how they can arise. The open-loop (steering wheel fixed) and closed-loop gust response dynamics were computed. Step and pulse crosswind inputs were modeled, and these were used to compute transient time responses of the driver/vehicle system both open- and closed-loop.

(4) Steering system dynamic lags and their significance to the closed-loop characteristics of the combined driver/vehicle system. Preliminary data were cited to show that steering lags may be as large as the driver's effective time delay. Because these delays are additive, the bandwidth and performance of the driver/vehicle system will be reduced accordingly.

The approach to these problems involved examining good driver/vehicle control structures to see ways in which they might be degraded and looking at poor situations to see how they might be improved or avoided. Many of the control concepts used (such as the adverse effect of opening an inner loop) are not new to a controls engineer, but it is their application to understanding the guidance and control of automobiles that represents an innovation. The examples shown, plus general application of the derived model to problems of this sort, provide the bases for remedial action that can help to avoid control difficulties (and reduce the hazard) or to enhance the guidance and control situation and increase traffic throughout.

APPENDIX

DRIVER/VEHICLE MULTILoop CLOSURES.—Multiloop closures of the driver/vehicle system are appropriate to providing good command following and good regulatory performance with minimum driver effort and good subjective opinion. Three different good multiloop structures have been evolved:

- (1) Inertial lateral deviation advanced in time
- (2) Path angle plus inertial lateral deviation
- (3) Heading angle plus inertial lateral deviation

Technical details of the closed-loop analyses needed to support table 5 of the main text are presented below, and the account is necessarily brief.

Inertial Lateral Deviation Advanced in Time.—This structure assumes that the lateral deviation along the path or track down the road can be perceived as shown in figure 6. The range to the point of regard is given by R , which is approximately equal to velocity U_0 times the time to travel to point T for small angles. Similarly, $R \sin \gamma$ is approximated by $R\gamma$.

The corresponding block diagram is given in figure 5a. The basic vehicle dynamics are given by the $y_I \rightarrow \delta_w$ transfer function. The time advance is obtained by differentiating y_I to obtain \dot{y}_I and then multiplying this by time T to the point of regard. The driver operates on the projected deviation to produce front wheel steer angle δ_w . The effective controlled element dynamics are given by

$$\mathcal{L} \left[\frac{y_I(t + T)}{\delta_w} \right] = (Ts + 1) \frac{y_I(s)}{\delta_w} \quad (17)$$

where \mathcal{L} denotes the Laplace transform. The effect of the perceptual preview is to provide pure lead term $(Ts + 1)$ in the controlled element that offsets the K_C/s^2 form of y_I/δ_w at low frequency. The amount of lead is given by R/U_0 , which is under the direct control of the driver and depends simply on how far down the road he is looking.

The basic $y_I \rightarrow \delta_w$ transfer function at 60 mph (88 fps) is obtained by double integration of the lateral acceleration in table 1, i.e.:

$$\frac{y_I}{\delta_w} = \frac{180[s^2 + 2(0.160)(7.24)s + (7.24)^2][s^2 + 2(0.293)(6.91)s + (6.91)^2]}{0.684s^2[s^2 + 2(0.785)(5.84)s + (5.84)^2][s^2 + 2(0.290)(6.97)s + (6.97)^2]} \quad 18)$$

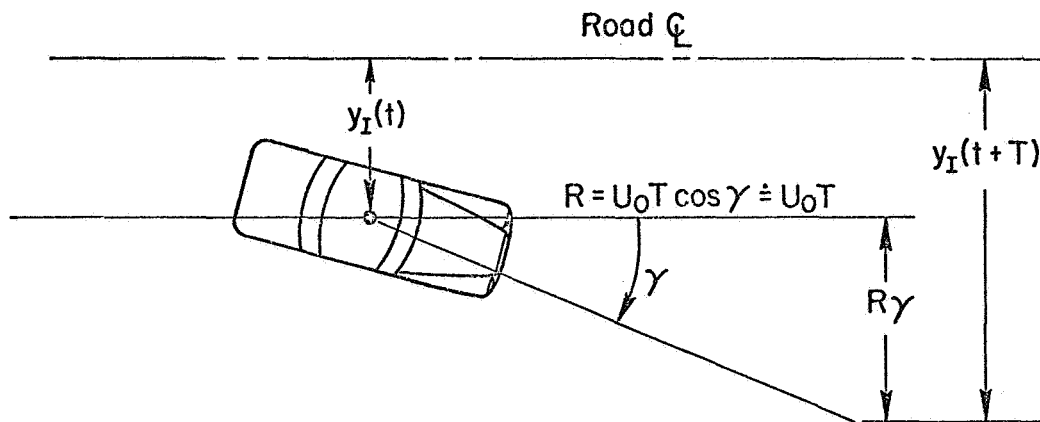


Figure 6.—Geometry for time-advanced lateral deviation.

A family of $j\omega$ Bode plots is given in figure 7 for the following values of perceptual lead:

$$T = 1 \text{ sec} = 88 \text{ ft at } 60 \text{ mph}$$

$$T = 2 \text{ sec} = 176 \text{ ft at } 60 \text{ mph}$$

$$T = 5 \text{ sec} = 440 \text{ ft at } 60 \text{ mph}$$

$$T = 10 \text{ sec} = 880 \text{ ft at } 60 \text{ mph}$$

A driver time delay of 0.3 second has been included in the phase plots.

The perceptual lead creates the desirable long stretch of nearly -20 dB/decade amplitude ratio slope in figure 7. For $T=10$ seconds a crossover at 1.2 rad/sec can be made with a 50° phase margin criterion (as opposed to the 60° criterion used in a later discussion). Values of T lower than about 5 seconds are not as good because they do not adequately compensate for the driver's time delay, and a good stable closure is harder to achieve. Values of T greater than about 10 seconds provide little added benefit under normal circumstances.

This multiloop structure is fine for regulation of lateral position on straight roads. Analytical difficulties arise when the model is asked to follow a command input such as curving roadway or a desired passing trajectory. As presented in figure 5a, it is necessary to construct an appropriately advanced command input via a guidance law and then to delay the driver's response by approximately the amount of the time advance. This delays the steering response of the driver/vehicle system until the previewed change in the path is reached.

Path Angle Plus Inertial Lateral Deviation.—This structure assumes that the inertial lateral deviation of the vehicle at its present position and the vehicle path angle are used as separate motion quantities. It differs from the time-advanced lateral deviation model discussed above by having the present lateral deviation available to the driver for comparison with the desired command input and derivation of a motion error. Time advance and delay guidance computations are not required. The two are indistinguishable if a command input is not present and if loop gain and preview are made compatible.

The block diagram for path angle plus lateral deviation control is given in figure 5b. A parallel structure is shown with the output of two driver blocks summing to give the net steer angle. A series structure could be employed as an alternative, with the output of the

driver's y_I control block providing an effective command at the input side of his γ control block. The differences are minor when the driver equalizations are simple gains plus a common time delay. The vehicle dynamics are given in the usual way by the transfer functions of reference 1.

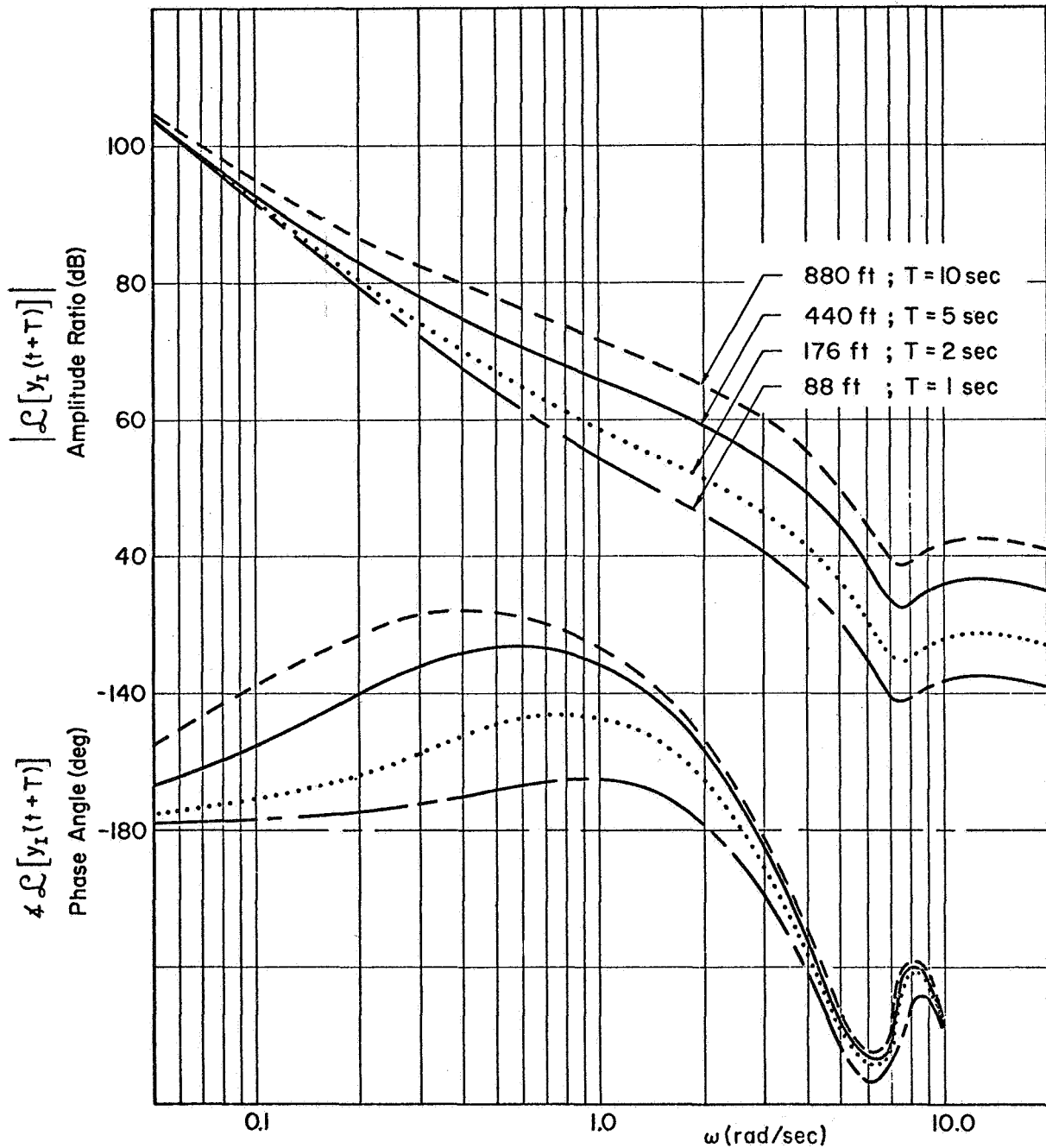


Figure 7.—Effective $y_I(t + T) \rightarrow \delta_W$ transfer function.

Assuming that the driver/vehicle task is to follow a desired passing trajectory or road-way curvature (denoted by y_{Ic}), the analytical problem is to determine the characteristics of the closed-loop driver/vehicle system. The formal procedure is to close the path angle loop and then to close the lateral deviation loop to obtain the $y_I \rightarrow y_{Ic}$ transfer function. The closed-loop function is:

$$\left[\frac{y_I}{y_{Ic}} \right]_{\substack{\gamma \rightarrow \delta_w \\ y_I \rightarrow \delta_w}} = \frac{Y_{py_I} N_{\delta_w}^{y_I}}{\Delta + Y_{p\gamma} N_{\delta_w}^{\gamma} + Y_{py_I} N_{\delta_w}^{y_I}} \quad (19)$$

where

$N_{\delta_w}^{y_I}$ is the lateral deviation numerator

$N_{\delta_w}^{\gamma}$ is the path angle numerator

Δ is the characteristic denominator

the vehicle numerators and denominator are known [i.e., ref. 1], and the driver describing functions Y_{py_I} and $Y_{p\gamma}$ are to be determined. Primes are used to denote an inner-loop closure and simplify the notation, i.e.,

$$\begin{aligned} \left[\frac{y_I}{y_{Ic}} \right]_{\substack{\gamma \rightarrow \delta_w \\ y_I \rightarrow \delta_w}} &= \left[\frac{y_I}{y_{Ic}} \right]'' \\ &= \frac{Y_{py_I} N_{\delta_w}^{y_I}}{\Delta' + Y_{py_I} N_{\delta_w}^{y_I}} \\ &= \frac{Y_{py_I} N_{\delta_w}^{y_I}}{\Delta''} \end{aligned} \quad (20)$$

The path angle (alone) system survey is given in figure 8, and this plot provides the inner-loop closure. No driver lead equalization and a time delay of 0.3 second will be assumed to provide a conservative estimate. A first-order Padé approximation to the reaction time delay is used; i.e., $e^{-Ts} \doteq -(s-2/\tau)/(s+2/\tau)$. The driver can cross over at about 1 rad/sec on the $j\omega$ Bode of figure 8 with good stability margin, giving a driver characteristic in the path angle loop of

$$Y_{p\gamma} \doteq 0.229e^{-0.3s} \doteq \frac{-0.229(s - 6.67)}{(s + 6.67)} \quad (21)$$

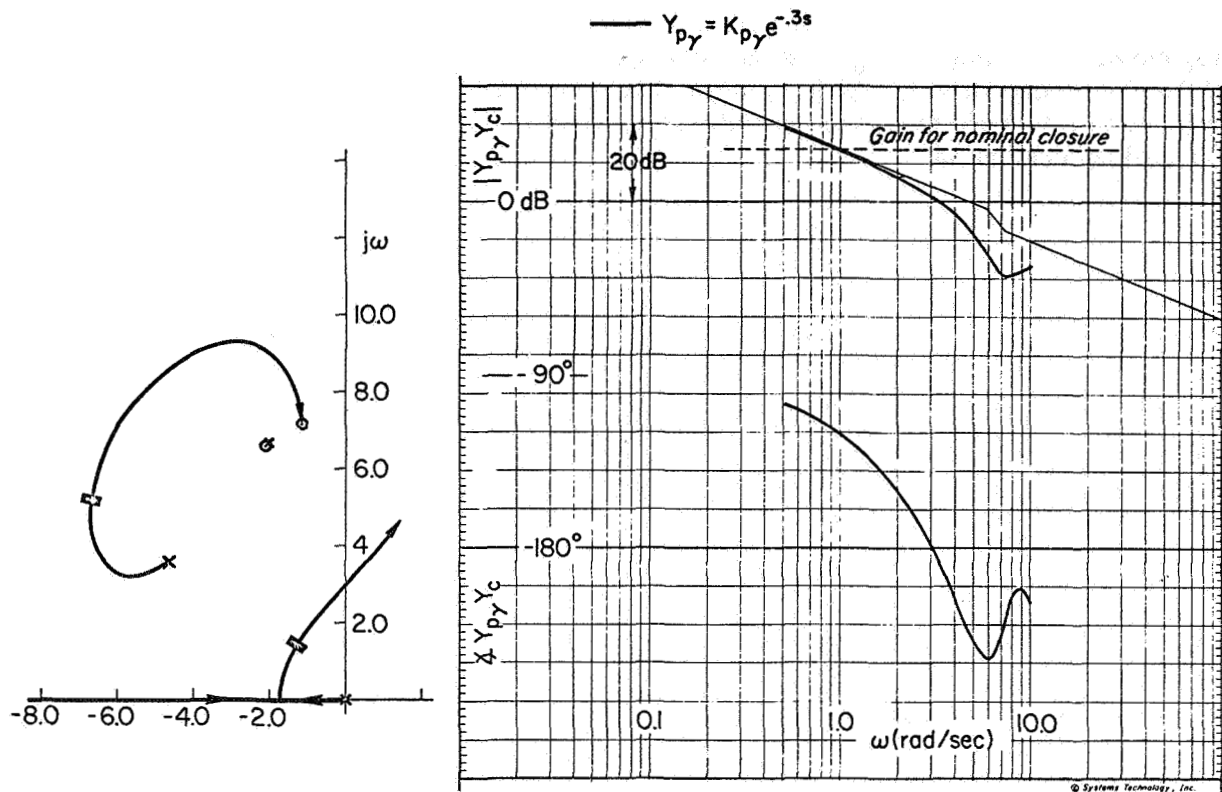


Figure 8.—Path angle loop closure.

This provides a good stable single-loop closure, and a higher gain might have been used in its role as an inner loop which would result in a higher outer-loop bandwidth. The units of equation (21) are radians of front wheel steer angle per radian of path angle error. In the notation of equation (20), the closed inner-loop denominator becomes:

$$\Delta' = \frac{.684 [s^2 + 2(0.639)(1.93)s + (1.93)^2] [s^2 + 2(0.290)(6.97)s + (6.97)^2] [s^2 + 2(0.798)(7.95)s + (7.95)^2]}{s(s + 6.67)} \quad (22)$$

The zero at -6.67 in the denominator of equation (22) results from the Padé approximation to the time delay.

The open outer-loop transfer function at 60 mph is given by:

$$\left[\frac{Y_I}{\delta_w} \right]' = \frac{180 \cancel{s} (s + 6.67) [s^2 + 2(0.293)(6.91)s + (6.91)^2] [s^2 + 2(0.16)(7.24)s + (7.24)^2]}{.684 \cancel{s} [s^2 + 2(0.639)(1.93)s + (1.93)^2] [s^2 + 2(0.290)(6.97)s + (6.97)^2] [s^2 + 2(0.789)(7.95)s + (7.95)^2]} \quad (23)$$

The Bode plot for equation (23) including the driver's outer-loop time delay (assumed 0.3 second) is given in figure 9. The crossover for 60° phase margin occurs at 0.53 rad/sec and that for neutral stability at about 1.5 rad/sec. A higher bandwidth in the outer loop could have been achieved by using a higher inner-loop gain.

This combination of feedback loops provides a system with adequate stability and command-following performance. The lack of driver lead equalization means that control

will not be sensitive to the driver's adaptation nor require his close attention, and the subjective opinion should be relatively good.

HEADING ANGLE PLUS INERTIAL LATERAL DEVIATION.—The driver is assumed to perceive vehicle heading angle and inertial lateral deviation at its present position as separate motion quantities. This structure differs from the preceding multiloop closure in that it uses only heading angle for the inner loop, not heading plus sideslip which gave path angle.

The block diagram for heading plus lateral deviation control is given in figure 5c. A parallel structure for the driver is used, although a series form could be employed with only the minor differences noted previously. The vehicle dynamics are given by the transfer functions of the vehicle at 60 mph.

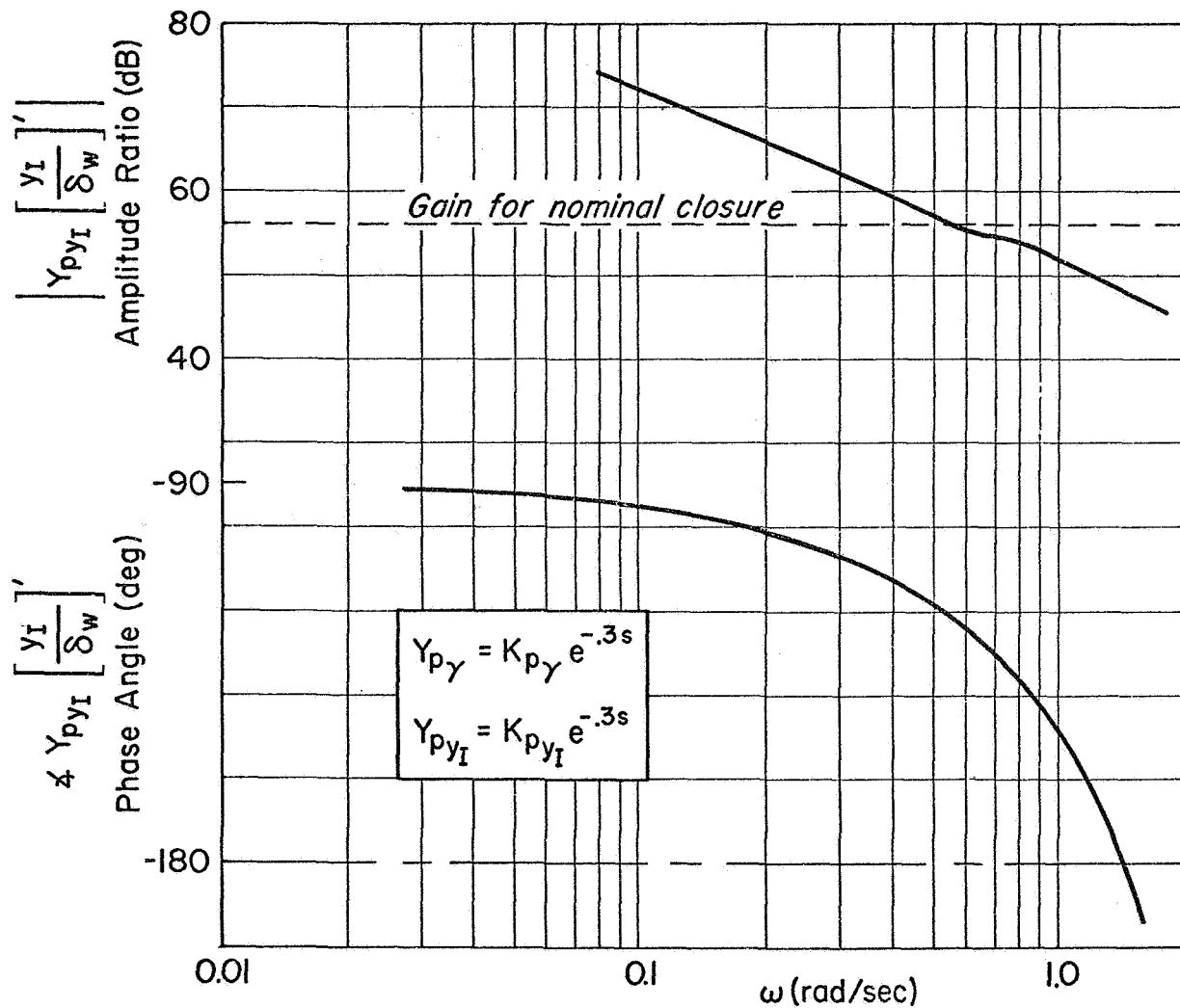


Figure 9.—Path angle plus lateral deviation closure.

The task is assumed to be the usual one of following a desired path or passing trajectory, defined as a lateral deviation command input. The closed-loop characteristics of the driver/vehicle system are to be determined by closing the heading and lateral deviation loops, successively.

The closed-loop transfer function is:

$$\begin{aligned} \left[\frac{y_I}{y_{Ic}} \right]'' &= \frac{Y_{py_I} N_{\delta_w}^{y_I}}{\Delta + Y_{p\psi} N_{\delta_w}^{\psi} + Y_{py_I} N_{\delta_w}^{\psi}} \\ &= \frac{Y_{py_I} N_{\delta_w}^{y_I}}{\Delta''} \end{aligned} \quad (24)$$

where $N_{\delta_w}^{\psi}$ is the heading angle numerator and the notation is that of equation (20).

The system survey for heading angle alone is given in figure 10, and this provides the basis for closing the inner loop in this multiloop case. The driver is assumed to have a pure gain plus time delay (0.3 second) form. A conservative crossover with good stability margins is 1.5 rad/sec, giving a driver heading response characteristic of:

$$Y_{p\psi} = 0.32e^{-0.3s} = \frac{-0.32(s-6.67)}{(s+6.67)} \quad (25)$$

The units of equation (25) are radians of front wheel steer angle per radian of heading error. The closed inner-loop denominator becomes:

$$\Delta' = \frac{.684(s+2.664)(s+9.397)[s^2+2(0.524)(3.61)s+(3.61)^2][s^2+2(0.289)(6.99)s+(6.99)^2]}{s(s+6.67)} \quad (26)$$

The zero at -6.67 in the denominator of equation (26) results from the Padé approximation, as before.

The open outer-loop transfer function for lateral deviation at 60 mph is:

$$\frac{y_I}{\delta_w} = \frac{180 \cancel{s} (s-6.67)[s^2+2(0.293)(6.91)s+(6.91)^2][s^2+2(.16)(7.24)s+(7.24)^2]}{0.684 \cancel{s} (s+2.664)(s+9.397)[s^2+2(0.524)(3.61)s+(3.61)^2][s^2+2(0.289)(6.99)s+(6.99)^2]} \quad (27)$$

The Bode plot for the effective outer-loop vehicle of equation (27) plus the driver's time delay (0.3 second) in the outer loop is given in figure 11. In addition to the nominal heading loop gain ($K_{p\psi} = 0.32$) case, the $(y_I \rightarrow \delta_w)'$ transfer functions are shown for several other heading loop gains. The nominal gain gives a lateral deviation loop crossover at 0.6 rad/sec with 60° phase margin. A slightly higher gain ($K_{p\psi} = 0.372$) gives a slightly better outer-loop system (crossover at 0.68 rad/sec) with adequate gain margin and the same phase margin. The differences are minor, however, and the nominal case was chosen to provide comparisons among the various loop structures with the same closure criteria.

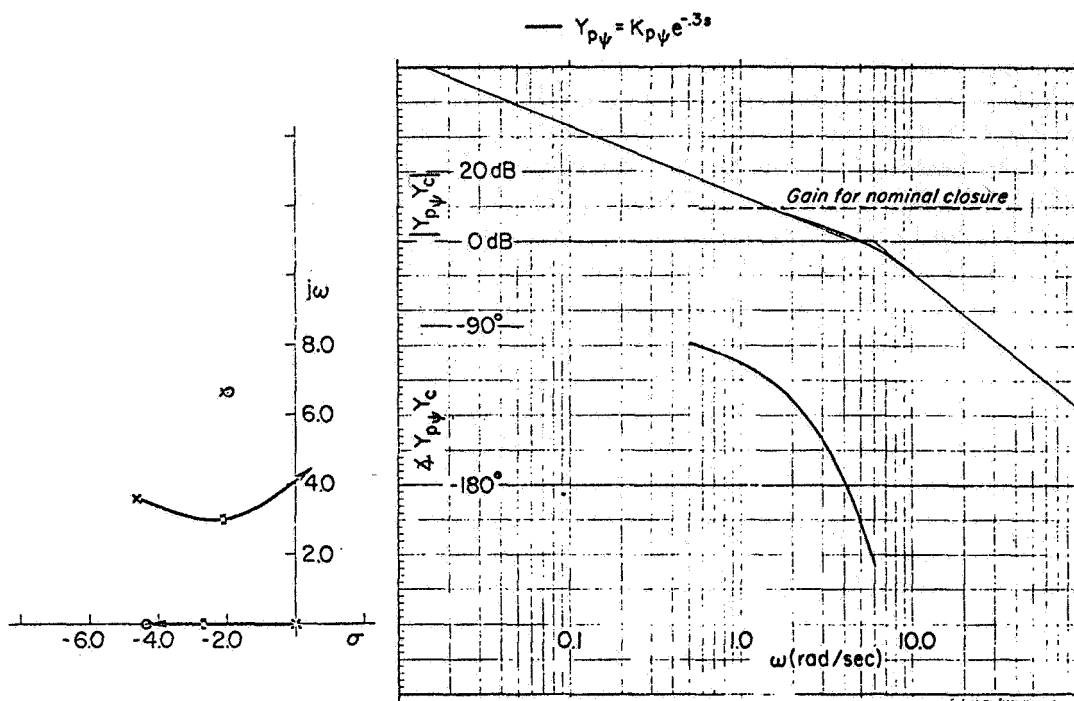


Figure 10. —Heading angle loop closure.

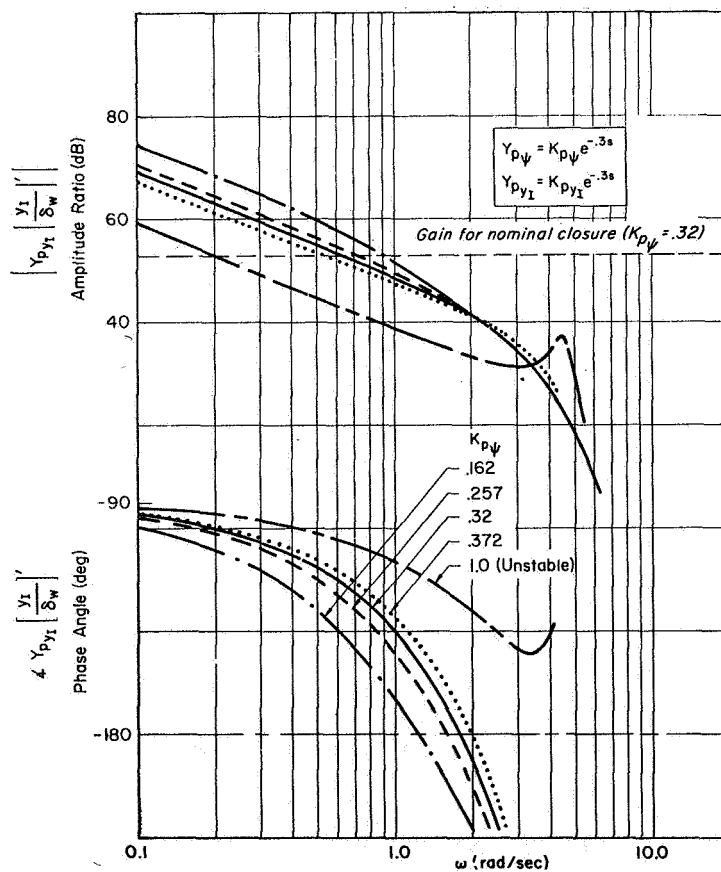


Figure 11. —Heading angle plus lateral deviation closure.

A large increase in heading loop gain results in a poor (and ultimately unacceptable) outer-loop system. This is illustrated in figure 11 for $K_{p\psi} = 1.0$, approximately three times the nominal value. This value of gain makes the low frequency quadratic pair in Δ' unstable (it has a negative damping ratio), and closing the outer loop simply makes this pair more unstable. The peaking of the outer-loop amplitude ratio due to a high inner-loop gain is illustrated by the $K_{p\psi} = 1.0$ case. Even if the damping were positive (but small, say, $K_{p\psi} = 0.9$) the crossover would have to be at a relatively low frequency to leave at least a 6-decibel gain margin between the top of the peak and the loop closing, the 0-decibel line.

REFERENCES

1. Weir, D. H.; Shortwell, C. P.; and Johnson, W. A.: Dynamics of the Automobile Related to Driver Control. Systems Technology, Inc., Tech. Rept. 157-1, July 1966.
2. Weir, David H.: Closed-Loop Directional Control of Automobiles. Paper presented at the IEEE Eighth Annual Symposium on Human Factors in Electronics, Palo Alto, Calif., May 3-5, 1967.
3. Weir, D. H.; and McRuer, D. T.: Driver Control During Overtaking and Passing. Systems Technology, Inc., Tech. Rept. 157-2, Apr. 1967.
4. Chandler, Robert E; Herman, Robert; and Montroll, Elliott W.: Traffic Dynamics: Studies in Car Following. Operations Research, vol. 6, no. 2, Mar.-Apr. 1958, pp. 165-184.
5. Todoseiv, Ernest Peter: The Action Point Model of the Driver-Vehicle System. Ohio State Univ., Eng. Exper. Sta., Rept. 202A-3, Aug. 1963.
6. Rothery, R.; Silver, R.; Herman, R.; and Torner, C.: Analysis of Experiments on Single-Lane Bus Flow. Operations Research, vol. 12, no. 6, Nov.-Dec. 1964, pp. 913-933.
7. English, John Jude: The Linear Mode of a System for the Automatic Control of Displacement of Vehicles. Ohio State Univ., Eng. Exp. Sta., Rept. 202A-4, Dec. 1963.
8. Herman, Robert; and Gardels, Keith: Vehicular Traffic Flow. Sci. Am., vol. 209, no. 6, Dec. 1963, pp. 35-43.
9. Salvatore, Santo: Vehicle Speed Estimation from Visual Stimuli. Public Roads, vol. 34, no. 6, Feb. 1967, pp. 128-131.
10. Fenton, Robert E.: Asymptotic Stability Studies in Simulated Car Following. Paper presented at Third Annual NASA-University Conference on Manual Control, Los Angeles, Mar. 1-3, 1967.
11. McRuer, Duane T.; and Krendel, Ezra S.: The Human Operator as a Servo System Element. J. Franklin Inst., vol. 267, no. 5, May 1969, pp. 381-403; no. 6, June 1959, pp. 511-536.
12. Wasicko, R. J.; McRuer, D. T.; and Magdaleno, R. E.: Human Pilot Dynamic Response in Single-Loop Systems with Compensatory and Pursuit Displays. AFFDL-TR-66-137, Dec. 1966.
13. Milliken, W. F.; Whitcomb, D. W.; Segel, L.; et al.: Research in Automobile Stability and Control and in Tyre Performance. Automobile Div., Institution of Mechanical Engineers, reprinted for Cornell Aero. Lab., Cornell Univ., 1956.
14. McRuer, D. T.; Ashkenas, I. L. and Pass, H. R.: Analysis of Multiloop Vehicular Control Systems. ASD-TDR-62-1014, Mar. 1964.
15. Segel, Leonard: On the Lateral Stability and Control of the Automobile as Influenced by the Dynamics of the Steering System. ASME Paper 65-WA/MD-2, Nov. 1965.

16. McRuer, Duane; Graham, Dunstan; Krendel, Ezra; and Reisener, Jr., William: Human Pilot Dynamics in Compensatory Systems—Theory, Models, and Experiments with Controlled Element and Forcing Function Variables. AFFDL-TR-65-15, July 1965.
17. McRuer, Duane T.; Graham, Dunstan; and Krendel, Ezra S.: Manual Control of Single-Loop Systems. J. Franklin Inst., vol. 238, no. 1, Jan. 1967, pp. 1-29; Feb. 1967, pp. 145-168.
18. McRuer, Duane: Remarks on Some Neuromuscular Subsystem Dynamics. IEEE Trans. vol. HFE-7, no. 3, Sept. 1966, pp. 129-130.
19. Stapleford, R. L.; McRuer, D. T.; and Magdaleno, R.: Pilot Describing Function Measurements in a Multiloop Task. NASA CR-542, Aug. 1966.
20. Weir, David H.; and Phatak, Anil V.: Model of Human Operator Response to Step Transitions in Controlled Element Dynamics. NASA CR-671, Jan. 1967.
21. McRuer, Duane T.; and Krendel, Ezra S.: Dynamic Response of Human Operators. WADC-TR-56-524, Oct. 1957.
22. McRuer, D. T.; Hofmann, L. G.; Jex, H. R.; et. al.: New Approaches to Human-Pilot/Vehicle Dynamic Analysis. Systems Technology, Inc., Tech. Rept. 164-2, June 1967.
23. Ashkenas, I. L.: A Consolidation of Lateral-Directional Handling Qualities. AIAA Paper 65-314, July 1965.
24. Meiry, Jacob L.: The Vestibular System and Human Dynamic Space Orientation. MIT Man-Machine Control Lab. Thesis T-65-1, June 1965.
25. Gordon, Donald A.: Experimental Isolation of Drivers' Visual Input. Public Roads, vol. 33, no. 12, Feb. 1966, pp. 266-273.
26. Gordon, Donald A.: Perceptual Basis of Vehicular Guidance. Public Roads, vol. 34, no. 3, Aug. 1966, pp. 53-68.
27. Calvert, E. S.: Visual Judgments in Motion. J. Inst. Navigation, vol. 7, no. 3, 1957, pp. 233-251.
28. Gibson, James J.: The Perception of the Visual World. Houghton Mifflin, Boston, 1950.
29. Schmidt, Ingeborg; and Connolly, Paul L.: Visual Considerations of Man, the Vehicle, and the Highway. SAE SP-279, Mar. 1966.
30. Biggs, Norman L.: Directional Guidance of Motor Vehicles—A Preliminary Survey and Analysis. Ergonomics, vol. 9, no. 3, May 1966, pp. 193-202.

10. Man-Machine Models for Car Steering*

*E. R. F. W. Crossman and H. Szostak
University of California*

The driver steering task and its environment are briefly described and the known properties of driver, highway, and vehicle reviewed. A three-level, information-processing model for driver steering control is developed.

Level 1 uses preview estimates of highway curvature as a basis for open-loop adjustment of vehicle path curvature. Level 2 uses relative heading, relative path-angle, or lateral position-rate to null out residual lateral position-rate errors. Level 3 uses lateral position itself in a low-frequency, low-gain loop to remove residual error. Evidence for each of these is briefly reviewed.

An auxiliary routine for controlling vehicle velocity to meet constraints on lateral acceleration is proposed as an adjunct to the model.

INTRODUCTION

This paper discusses the structure of what is almost certainly the most common manual control task in modern life, viz. car steering. Back-of-the-envelope calculations indicate that some 10^{10} to 10^{11} manhours are spent per annum driving automobiles; a major part of this task is keeping them in lane while traveling on highways, maneuvering them in and out of parking lots, etc. Since the penalties for failure are high including damage to property, personal injury and lost time, extreme reliability is demanded of the man-machine system. It is in general achieved, as the low accident-rates per vehicle mile indicate. This fairly satisfactory position appears to have been reached by evolutionary improvement in vehicle and highway design and maintenance, aided by the human driver's ready acquisition of control skills and his great adaptive powers, rather than by an explicit analytical design procedure, and in this respect car driving resembles most other manual control systems.

The residual failure rate which is currently a matter of public concern, must be associated with faults or breakdowns in one or more of the three components: driver, vehicle, and highway. The fact that there are numerous distinct types of failure, each making a small contribution to the overall rate, indicates that the evolutionary process must have reached a steady state or plateau. Unless revolutionary changes such as automatic

*The research for this paper was supported by the Institute of Urban and Industrial Health, U.S. Public Health Service, under Grants Ac-00260 and UI-00016.

steering are introduced, further progress can only be sought through attention to numerous relatively small details and specific features of the man-machine system, each by itself producing only a small increment of performance and reliability. To facilitate this process, a fully analytical and experimentally substantiated theory of the highway/driver/vehicle control system is required. The present paper sets out our present thinking on the form such a theory should take.

The general outline of a car steering theory has been developed by several previous authors, but at present there appears to be no general agreement either on exact topology or on quantitative estimates of the various parameters involved. Rashevsky (ref. 1) and Wohl (ref. 2) pointed out the relatively simple closed-loop structure of the steering task but did not give adequately detailed treatments nor cite experimental evidence in support of their analyses. Weir (ref. 3) has reviewed the possible loop-closures and performance likely to be obtained from them, but does not take preview and open-loop response modes into consideration. Perhaps the most explicit frequency domain model of the driver is that recently proposed by Wierwille et al. (ref. 4) derived from experimental model-matching studies of three subjects steering a winding-highway task at a fixed 68.2 mph under full (450 ft) and zero preview in a fixed base simulator with pure second-order vehicle dynamics ($0.185/s^2$). The model developed for the former condition represents the driver's output as the sum of a feedforward (open-loop) transfer from highway lateral displacement (a polynomial in the Laplace-transform variable s), and a feedback (closed-loop) transfer from vehicle lateral position error, referred to an arbitrary zero. While feedforward control was introduced for the first time, no attempt was made to separate the various feedforward and feedback paths that could in principle be employed by the driver using the different channels of visual, kinaesthetic, and vestibular information available to him, nor was the effect of forward speed on vehicle dynamics taken into account.

In the present paper we first discuss the known properties of the three major system components—highway, driver, and car—in a generalized way, and then present an outline model of the car/driver/highway interaction based on these and on certain of our recent experimental results. No attempt is made to develop a full analytical treatment, partly for reasons of space and partly because we feel that at the present stage discussion should be focused on the macrostructure and topology of the system before getting down to specifics of the analytical formulation. Following this line of thought we have mixed digital-type information processing models rather freely with continuous-control models, choosing whichever seems most suitable at each point. Once precisely formulated, each of these may be rewritten in terms of the other to yield a consistent overall treatment. But we do not yet know which is more useful.

ENVIRONMENT AND TASK DESCRIPTION

Consider the general cross-country driving situation. This includes both straight and winding roads, with and without traffic. Speeds range between zero and, say, 75 mph, and may at times fluctuate rapidly. The pavement may be level, tilted, or uneven, and there will be side-forces due to wind, both natural and that caused by other vehicles. The visual field will generally be rich in detail and structure; forward visibility is normally several hundred feet, but may be restricted by fixed obstructions, by other traffic, by darkness, and by atmospheric conditions such as rain, fog, and snow. In the latter case we may assume a 200-foot minimum for normal driving.

The steering task varies from moment to moment when passing and fixed obstacles are considered, but for a large part of the time the driver must stay within his traffic lane despite disturbances. With little loss of generality we may assume the lane defining the

lateral position tolerance to be 12 feet wide, demarcated by white painted lines on either side. If the car is 6 feet wide there will be no penalty associated with fluctuations in lateral position up to ± 3 feet from the center of this lane, but in certain frequent circumstances (presence of fixed obstacles and other traffic) the "cost" of errors exceeding this increases steeply.

The control system currently utilized may conveniently be split into three components: driver, vehicle, and highway geometry (fig. 1). The latter takes into account the consequences of forward motion and of the driver's distant vision. Since the properties of highway and vehicle and the geometrical relationships are all well understood, our current research problem is to model the driver adequately enough to account for the known facts of steering performance, and to permit predictions under variation of design parameters.

STEERING SYSTEM COMPONENTS

THE DRIVER. — Certain general remarks can be made about the driver's information processing capability which limit the scope of possible driver models.

(1) With his remarkable pattern recognition powers, the driver may be assumed capable of acquiring any desired signal or data present in his visual field, filtering and correcting distortion. However, the precision of any quantitative estimate cannot exceed around 3 bits per dimension per observation (ref. 5).

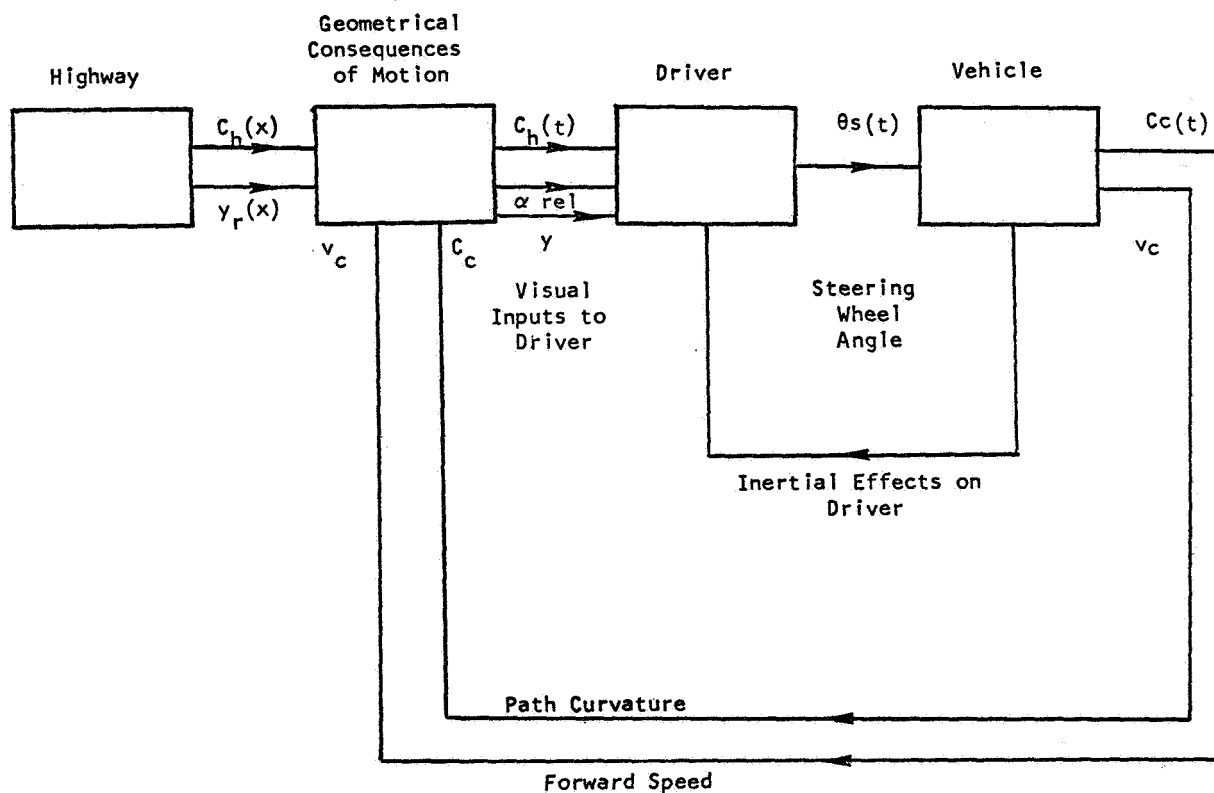


Figure 1. —General layout of the steering subsystem.

(2) The selective power of the driver's eye-fixation system and brain permits independent data to be acquired from visual fields as small as 3×10^{-4} or as large as 3×10^{-2} steradians (1° - 10° subtense). However shifts of attention take significant time so that frequent switching is uneconomical and attention is likely to be focused on only one area of the visual field at a time.

(3) The driver may take the first time derivative of current input signals, but this entails a marked loss of precision, and, it is safe to assume, that only 1 to 2 bits of rate information are available for any observed variable.

(4) Judgments of relative time are also subject to about a 3-bit limitation so that precise purposeful time delays or prediction cannot be assumed.

(5) Inputs from vestibular senses and body musculature provide both linear and angular acceleration data for, respectively, the head and the seat of the pants, but again with strictly limited precision (ref. 6).

(6) The effector subsystem, normally two hands applying torque to the steering wheel, can set an angular position or apply a desired torque with strictly limited precision (about 2-1/2 bits); it can hold an existing position against changes of torque; and produce a desired angular rate or torque rate with somewhat lower, perhaps 2-bit, precision.

(7) The effector subsystem may also supply data ("feel") on the torque currently being applied and on high-frequency load-changes or torque inputs.

(8) There is no evidence that central processes can operate at any higher precision than the input and output channels described above.

(9) There is a strong likelihood that central computations and decisions cannot be multiprocessed, hence that sequential programing must be assumed. However input and output buffers permit some flexibility of central process timing.

(10) Successful task performance normally requires a prolonged period of skill acquisition, hence we may assume complex information processing programs and the presence of much data in permanent storage.

Considering these constraints, the natural form of driver model would seem to be a sequential information processor like a digital computer, but with severely limited transient data storage, word-length 2 or 3 bits, and a small instruction set. Long sequences of processes would be substituted for precision. This viewpoint is supported by recent studies on principles of direct digital control (ref. 7).

If a quasi-linear continuous control model is adopted, the generally low precision of the driver's input data-channels and output adjustments and sequential computation capacity suggests that it must assume low-order dynamics except at frequencies directly involving neuromuscular dynamics, which are not generally of interest in car steering. We must also expect low correlation (coherency) between input and output.¹ Since noise is introduced in processing, we must expect the driver to combine data obtained through two or more channels to provide enough information to fully utilize his output capacity.

THE HIGHWAY. — The desired path of the vehicle along the highway, which is the primary disturbance or forcing-function input to the vehicle-driver system, can be best specified as a time-invariant space function $\alpha_h(\underline{x}_h)$ where

α_h = tangent to the direction of the highway centerline

\underline{x}_h = arc-length position along the highway

¹However if all the information in a single 3-bit input were reproduced at the output the correlation obtained would be $r=0.92$, and even 1-bit transfer yields $r=0.77$.

Since highways are generally constructed out of straight sections connected by circular arcs, the function $\alpha_h(x_h)$ is a special kind of stochastic variable, whose first derivative, the highway curvature $c_h = d\alpha_h/dx_h$ takes a sequence of randomly chosen constant values for successive sets of values of x_h . The topography of a given highway is most economically described by the joint distribution of curvatures c_h and distances between transition points Δx_h . To illustrate this point, tables 1 and 2 and figure 2 give the successive curvatures and arc-length distances between transition points for a test section of highway used in our experiments. c_h appears usually to have zero mean and a unimodal distribution with range on the order ± 0.2 deg/ft, while Δx_h has a positive mean and unimodal distribution. While highway data are readily available, their topography does not appear to have been subjected to statistical analysis, so that little more can be said of the highway-curvature forcing function.

A second group of highway-produced disturbances is that due to sidewinds, irregular pavement, etc. These seem to be best described collectively as a random zero-mean addition to the vehicles' yaw rate. Beyond the fact of their existence we have little data to go on, but it seems plausible to assume that the total forcing function, which is the sum of many small incoherent components, can be described as a Gaussian disturbance with limited bandwidth and variance dependent on forward speed.

Highway surface friction is also important since in combination with tire mechanics it provides side forces for vehicle steering. In general we assume high μ conditions providing a large linear range of steering response.

VEHICLE DYNAMICS.—As shown by Wohl (ref. 2) among others, a purely "kinematic" vehicle would have a linear transfer function K/s^2 from steering wheel angle to lateral position on the highway. This can be considered to be generated in two tandem stages

- (1) a time-independent effect of steering angle on path traversed
- (2) the effect of forward speed which transforms path traversed into a time function

This ideal second-order response is actually obtained at very low speeds. At higher speeds a variety of secondary dynamic effects appear. These have been studied by Segel (ref. 8), Nordeen (ref. 9), Weir, Shortwell, and Johnson (ref. 10) and many other authors under the general heading of vehicle handling characteristics. Space does not permit full discussion of the many theories and data presented, but certain generalizations may be plausibly asserted to hold true for most passenger cars.

(1) At limiting low forward speeds ($v_c = dx_c/dt \rightarrow 0$, where v_c = vehicle forward speed), the curvature c_c of the car's path in space is proportional to steering wheel angle θ_s . This is the ideal, Ackerman, steering response. Hence

$$c_c = \left(\frac{\delta \alpha_c}{\delta x_c} \right) = K_0 \theta_s \quad (v_c \rightarrow 0) \quad (1)$$

K_0 is typically about 6.5×10^{-3} deg/ft/deg. This point is illustrated by tables 3 and 4, which give the distribution of values of the ideal curvature-response gain K_0 derived from data for 93 European cars tested by the British Consumer's Association in the period 1962 to 1967 and reported in *Motoring Which?* (ref. 11).

Equation (1) can be rewritten in the small-signal form

$$\left[\frac{\partial c_c}{\partial \theta_s} \right]_{v \rightarrow 0} = \text{Const.} \quad (2)$$

TABLE 1.—TOPOGRAPHY OF BEAR CREEK ROAD,
CONTRA COSTA COUNTY, CALIF.

Segment	Length (feet)	Horizontal Curvature (deg/feet)	Segment	Length (feet)		Segment	Length (feet)	
1	302	0.000	21	1,476	0.000	51	627	0.000
2	1,696	+0.088	22	782	-0.064	52	1,129	+0.088
3	598	0.000	23	270	0.000	53	2,641	0.000
4	859	-0.088	24	273	+0.027	54	756	-0.082
5	2,003	0.000	25	272	0.000	55	55	0.000
6	344	-0.088	26	876	-0.042			
7	510	0.000	27	626	0.000			
8	639	+0.088	28	674	-0.057			
9	490	0.000	29	506	0.000			
10	539	-0.088	30	653	-0.088			
11	387	0.000	31	404	0.000			
12	1,118	+0.112	32	740	+0.052			
13	400	0.000	33	815	0.000			
14	400	-0.088	34	284	+0.088			
15	524	0.000	35	400	0.000			
16	465	+0.029	36	414	-0.088			
17	587	0.000	37	300	0.000			
18	640	-0.019	38	615	+0.088			
19	2,820	0.000	39	400	0.000			
20	1,027	-0.028	40	775	-0.088			
Total Length			33,211 feet					
Mean Length of Segments			736 feet					
Mean Curvature of Segments			0.01 degree/ft					

(2) Equation (2) continues to describe the small-signal low-frequency response of the vehicle throughout the working range of forward speed and curvature but the curvature-response gain parameter $K(v_c, c_c)$ is a weak function both of forward speed v_c and of mean path curvature c_c

$$\left[\frac{\partial c_c}{\partial \theta_s} \right]_{v_c, c_c} = K(v_c, c_c) \quad (|v_c| > 0) \quad (3)$$

TABLE 2.—DISTRIBUTION OF LENGTH AND CURVATURE
OF HIGHWAY SEGMENTS

Length	Number	Curvature	Number
0 - 199 feet	0	0.125 - 0.175 deg/feet	0
200 - 399 feet	8	0.075 - 0.124 deg/feet	6
400 - 599 feet	14	0.025 - 0.074 deg/feet	3
600 - 799 feet	11	-0.025 - 0.024 deg/feet	23
800 - 999 feet	3	-0.075 - -0.026 deg/feet	4
1000 - 1199 feet	3	-0.125 - -0.076 deg/feet	3
1200 - 1399 feet	0	0.175 - -0.126 deg/feet	0
1400 - 1599 feet	1	Total Segments	44
1600 - 1799 feet	1	Mean Curvature	.0056
1800 - 1999 feet	0	S.D.	.017
2000 and over feet	3	Variance	.003
Total Segments	44		
Mean Length	736 ft		
S.D.	167 ft		

The curvature-response gain function has not been fully explored in the literature. Published data of Segel (ref. 8) on a 1953 Buick sedan and our own measurements on a 1965 Ford sedan test car (ref. 12) (fig. 3(a)), indicate that at constant c_c there is a 60 or 70 percent reduction in K over the normal speed range (0 to 70 mph). However, published data for certain high-performance vehicles such as the Lotus Elan tested by Motoring Which? (ref. 11) show very small percentage decrease of gain with speed (fig. 3(b)). Varying c_c with constant v_c is a less frequently performed test, and results here show quite variable patterns (ref. 11). However, subjectively better handling cars again show small percentage changes in K over the normal curvature range $c = \mp 0.5$ deg/ft. Taking the available data together, we can reasonably infer that ideal car-handling qualities would call for completely linear curvature response, that is the small-signal path curvature response gain K should be invariant under changes of forward speed and of mean curvature. This response pattern can of course only be maintained up to some limiting value of lateral acceleration a_r generally between 0.6 to 0.9 g where

$$a_r = c_c \cdot v_c^2 \quad (4)$$

Above this point K begins to change rapidly either downwards (limiting understeer) or upwards (limiting oversteer) and even approximate linearity can no longer be assumed.

(3) The dynamic response of path curvature to steering wheel input is mechanically complicated but appears to be well approximated for frequencies of interest in normal steering by a second-order filter with breakpoint about 1.0 hertz. Here again there is a serious lack of data from which valid generalizations might be derived. Our own studies on a 1965 Ford

sedan test car (fig. 4) indicate that linear response is maintained up to perhaps 10 hertz with some high-order dynamics due mainly to tire and suspension effects. The main break-frequency around 1 hertz presumably varies with type of car, forward speed, and path curvature, but in an unknown manner.

Yaw angle and lateral position displacement responses can be readily derived geometrically from the path-curvature responses outlined above and are respectively proportional to v_c and v_c^2 .

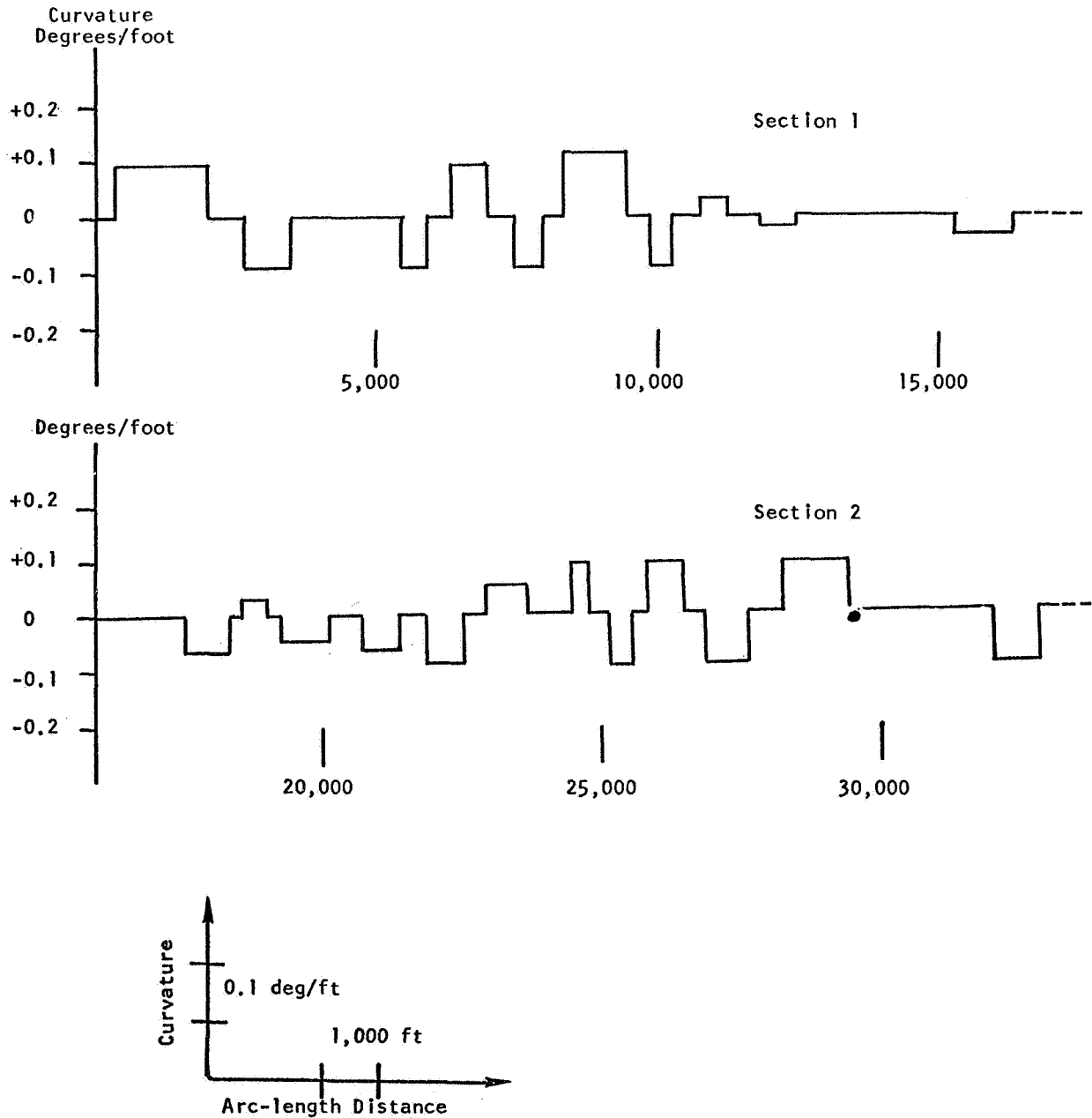


Figure 2(a). — The highway as forcing function. A 33 000-foot section of Bear Creek Road, Contra Costa County, Calif.

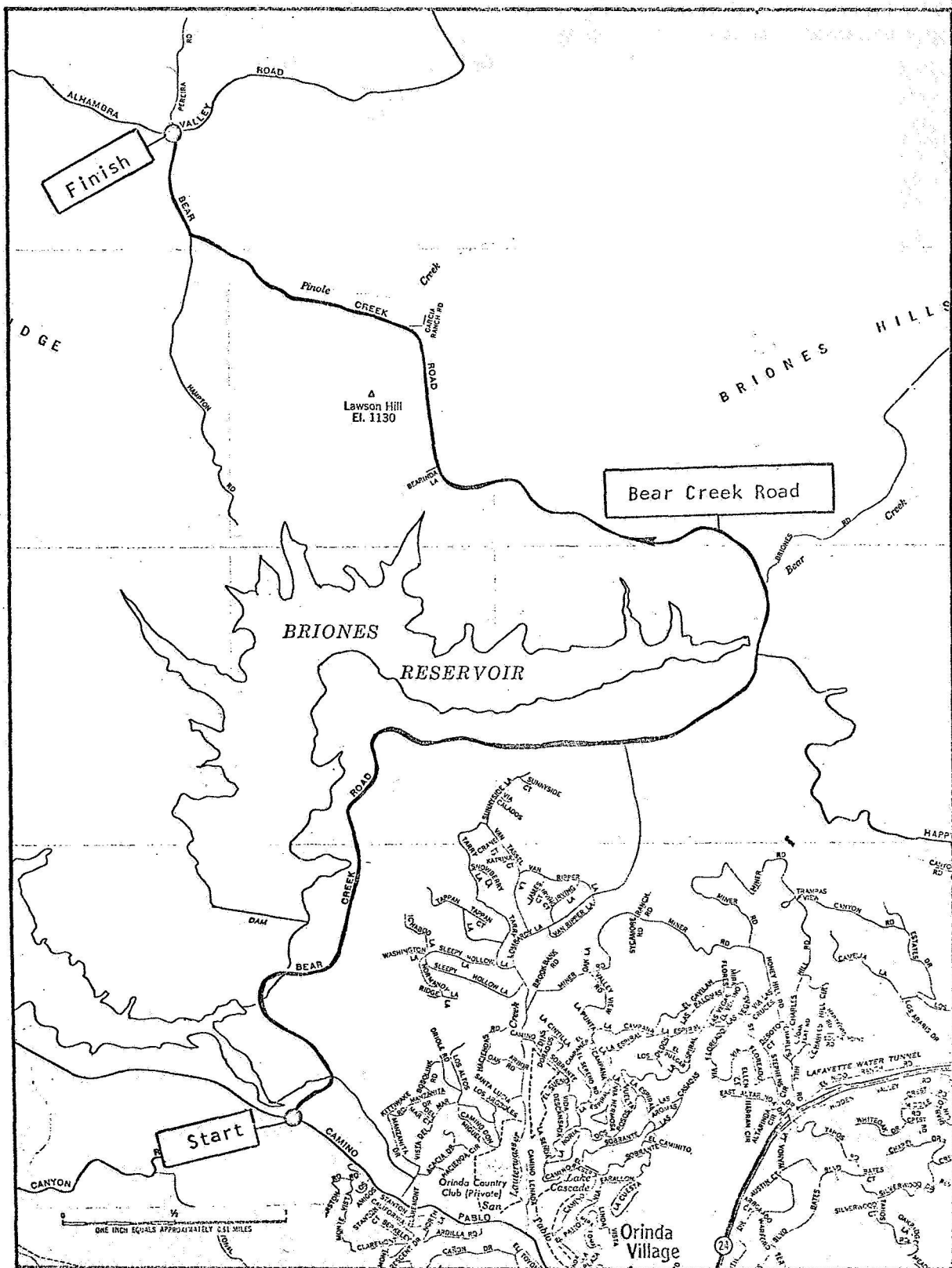


Figure 2(b). —Plan view of Bear Creek Road.

TABLE 3.—SMALL-SIGNAL CURVATURE-RESPONSE GAIN OF 100 EUROPEAN CARS
TESTED BY CONSUMER'S ASSOCIATION, LONDON, 1962-7

Gain (deg/ft/deg x 1000)	Frequency (Percentage)
3 - 3.9	0
4 - 4.9	4
5 - 5.9	29
6 - 6.9	22
7 - 7.9	17
8 - 8.9	10
9 - 9.9	0
10 and over	
Mean Gain	0.0065 deg/ft/deg
S.D.	0.0011 deg/ft/deg

TABLE 4.—PERCENT REDUCTION OF CURVATURE-RESPONSE GAIN AT 0.25 g
LATERAL ACCELERATION AT TWO CONSTANT CURVATURES FOR 51 CARS
MEASURED BY CONSUMER'S ASSOCIATION, LONDON

Percent Reduction	Frequency (Percentage) Curvature 0.76 deg/ft (75 ft radius)	Frequency (Percentage) Curvature 0.155 deg/ft (350 ft radius)
0 - 9	8	0
10 - 19	36	0
20 - 29	54	4
30 - 39	4	8
40 - 49	0	19
50 - 59		42
60 - 69		27
70 - 79		0
	n = 25	n = 26
Mean Percent Reduction	20.3	53.0

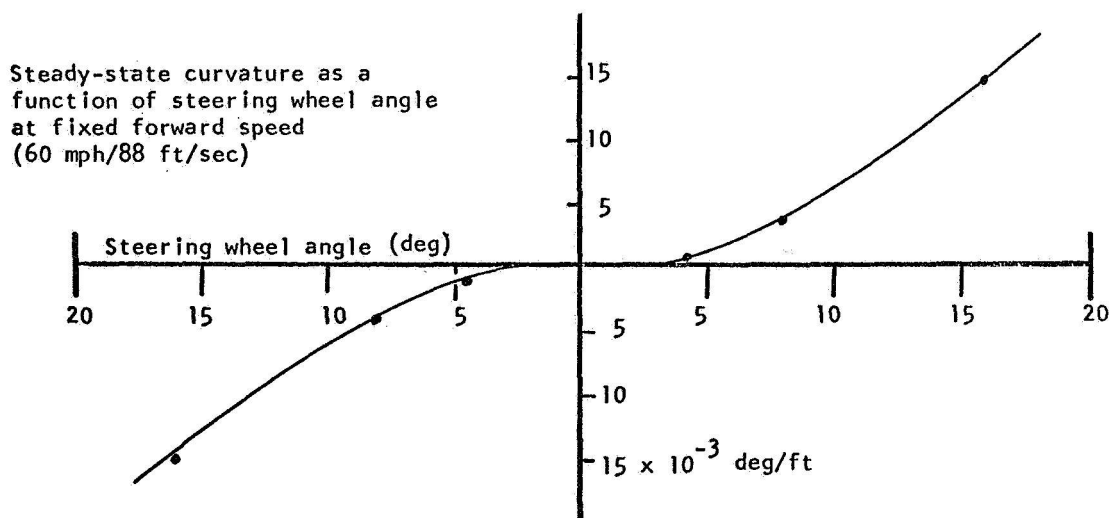
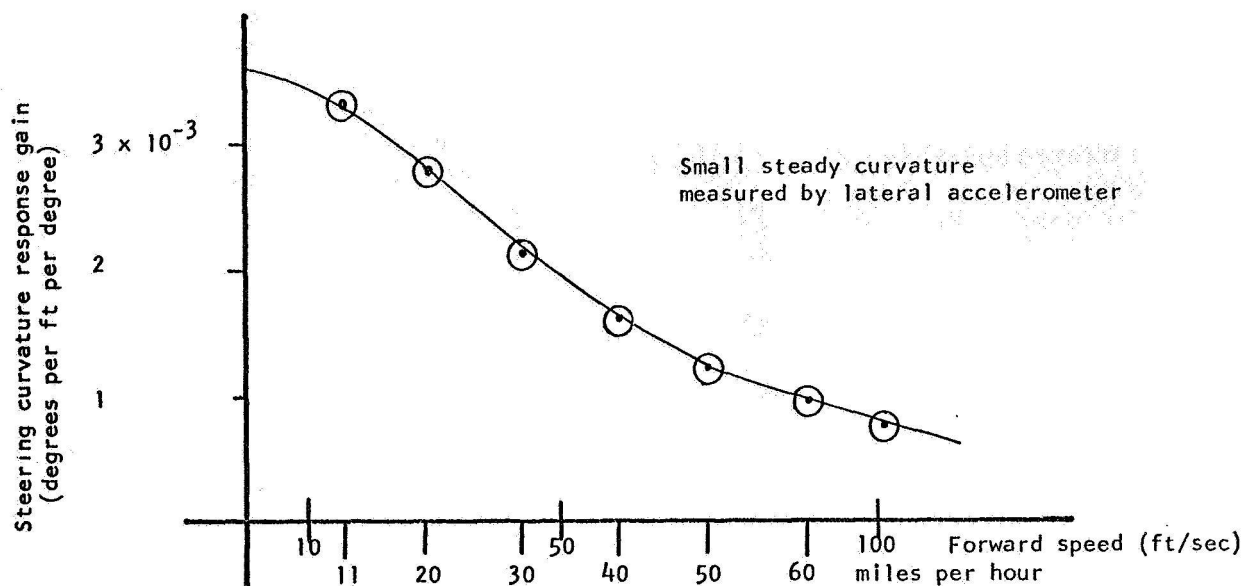


Figure 3(a).—Measured curvature response of a 1965 Ford sedan.

(4) Torque approximately proportional to angular displacement is needed to deflect the steering wheel from its central or straight-ahead position. Numerical values on the order of 0.1 lb/deg are typically recorded for non-power-steering cars (ref. 11).

A THREE-LEVEL DRIVER MODEL

LEVEL 1: OPEN-LOOP CURVATURE CONTROL USING PREVIEW.—Since highways tend to show long sections of constant curvature connected by well-defined transitions to new curvature values, and since the vehicle has approximately constant gain from steering wheel angle to path curvature, a minimal control policy could be implemented as follows: The

driver scans the road ahead and detects points at which the highway changes curvature. He estimates the new curvature while approaching the transition point, translates this mentally into a required new steering wheel setting, computes the torque to be applied, and executes the required change at about the time the car comes abreast of the transition point.

Given a car with ideal steering characteristics this policy could be executed independently of forward speed, thus decoupling the steering and speed control subsystems, a major advantage given the normal requirement for frequent speed-variations due to traffic, gradients etc. However, even in this case the success of the policy would obviously be limited by the precision of the driver's curvature estimation. No data are available from actual highway situations to estimate this parameter, but laboratory studies by Pollack (ref. 13) indicate about 2-bit precision. While this figure may well be subject to improvement under the kind of extended practice with knowledge of results that is obtained in normal driving, we should probably assume that less than 3.0 bits of absolute curvature information would be available,

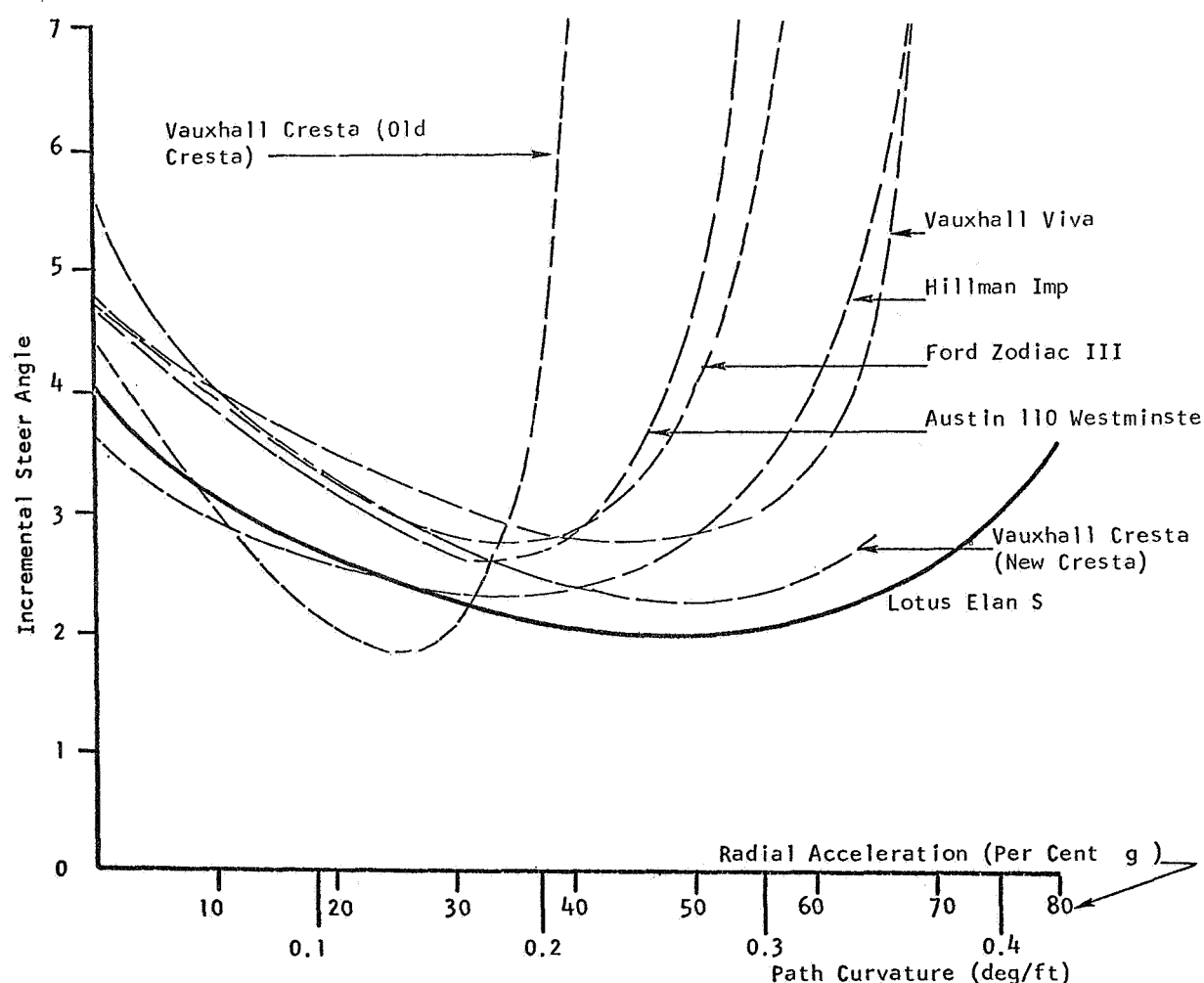


Figure 3(b).—Incremental steer angle per unit increase in curvature, measured by Consumers Association, London. Constant speed 40 mph, inward spiral course.

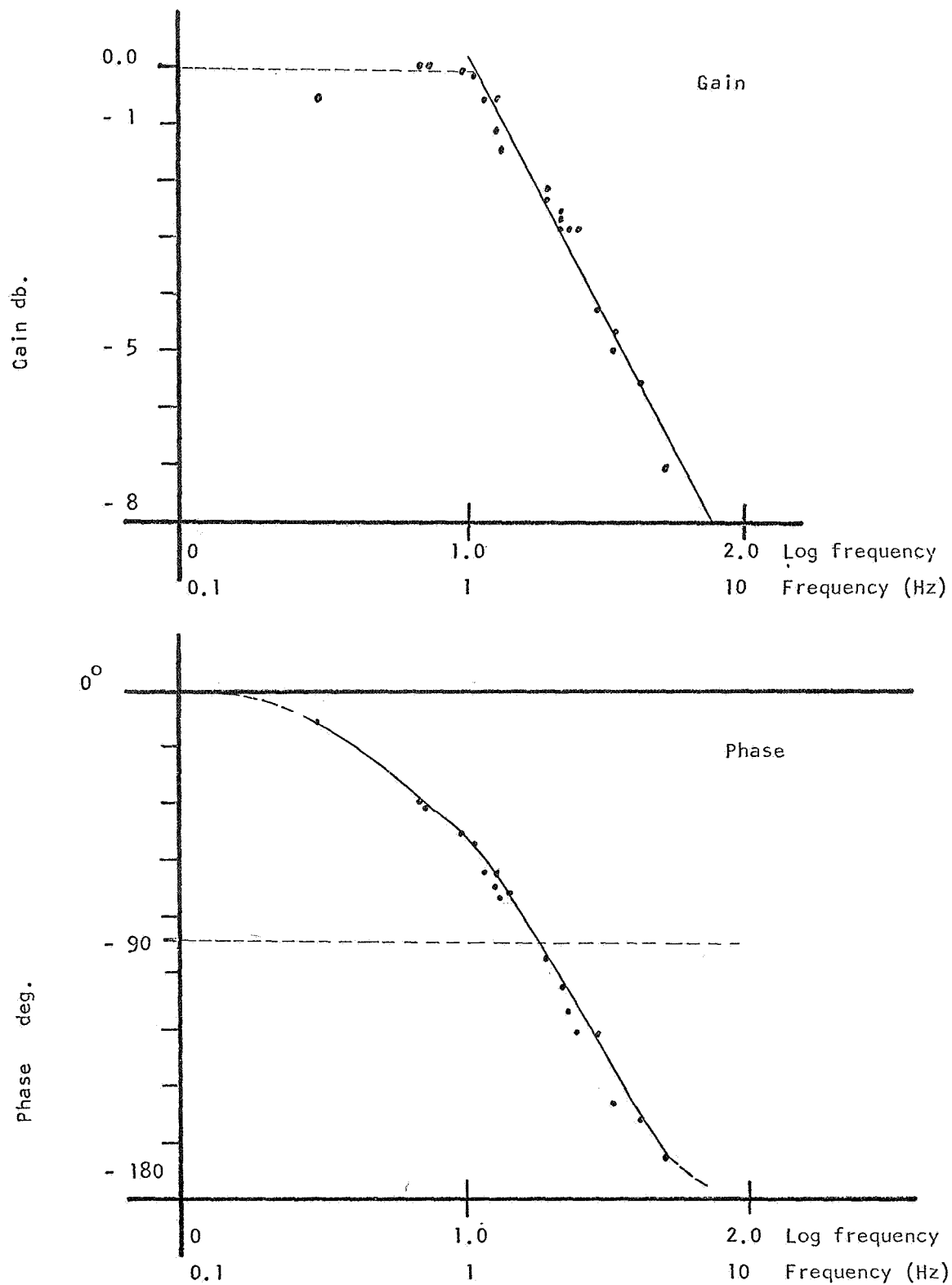


Figure 4.—Measured yaw-rate and path curvature frequency-response of a 1965 Ford sedan (ref. 12).

including the 1-bit sign. Since information would undoubtedly be lost in computing the required steering adjustment, this open-loop control model must be assumed capable of producing a vehicle path curvature correlated less than $r=0.90$ with highway curvature. Since lateral-position error is a double time-integral function of curvature error, this control mode is clearly incapable of fully meeting the performance criteria outlined above. However, its adoption would substantially reduce the load placed on the closed-loop modes described below.

Since this is a strictly open-loop policy it requires the output member to be in some sense calibrated to permit setting at a desired value. This appears to be accomplished by combined use of a central (zero-curvature) straight-ahead reference position obtained (in an ideal car and to good approximation in most real cars) by applying zero torque to the steering wheel, and an approximately linear torque/displacement curve. The driver may be supposed to apply a known torque to the wheel in the expectation that this will result in a proportional path curvature.

We have little direct support for this conjecture, but we have found it essential to provide our simulator with realistic feel including a central reference point. This is achieved by the use of opposed springs providing force proportional to angular displacement from the central straight-ahead position. Numerous steering wheel torque measurements reported in *Motoring Which?* (ref. 11) also support this view.

A suitable flow-chart for executing this hypothetical first-level control policy is shown in figure 5. Assuming the component processes to be computed on a timescale of fractions of a second, the corresponding continuous transfer-function driver model would be a simple proportionality with added noise, the cutoff frequency being considerably higher than that needed to respond adequately to highway variations (fig. 6).

The only direct data we can currently adduce to support this hypothesis are recordings of eye fixations (fig. 7) which show that the driver's attention is certainly placed where it would be best able to acquire the necessary data, viz. in the middle distance and deflected in the direction of curves currently being negotiated. Further detailed study of eye-fixation and steering response patterns will be needed, and we also plan to analyze the pattern of driver response to curvature transitions in search of the predicted pattern of advance preparation and transient error.

LEVEL 2: CLOSED-LOOP CONTROL OF RELATIVE HEADING AND/OR RELATIVE PATH ANGLE AND/OR LATERAL POSITION RATE.—Assuming that the above open-loop policy can only achieve partial success, there will be residual curvature error, the difference between highway and car-path curvature. Let this be c_e , where

$$c_e = c_h - c_c \quad (5)$$

At a given forward speed v_c this results geometrically in a proportional time rate of increase of the angle between car-path heading and highway direction, the relative path angle $\alpha_{rel} = \alpha_h - \alpha_c$

$$\frac{d(\alpha_{rel})}{dt} = \frac{d\alpha_h}{dx_c} \cdot \frac{dx_c}{dt} = v_c \cdot c_e \quad (6)$$

$$\frac{d^2 y_c}{dt^2} = \frac{d(\alpha_{rel})}{dt} \cdot \frac{dx_c}{dt} = v_c^2 \cdot c_e \quad (7)$$

where y_c = lateral position of the car relative to the highway centerline.

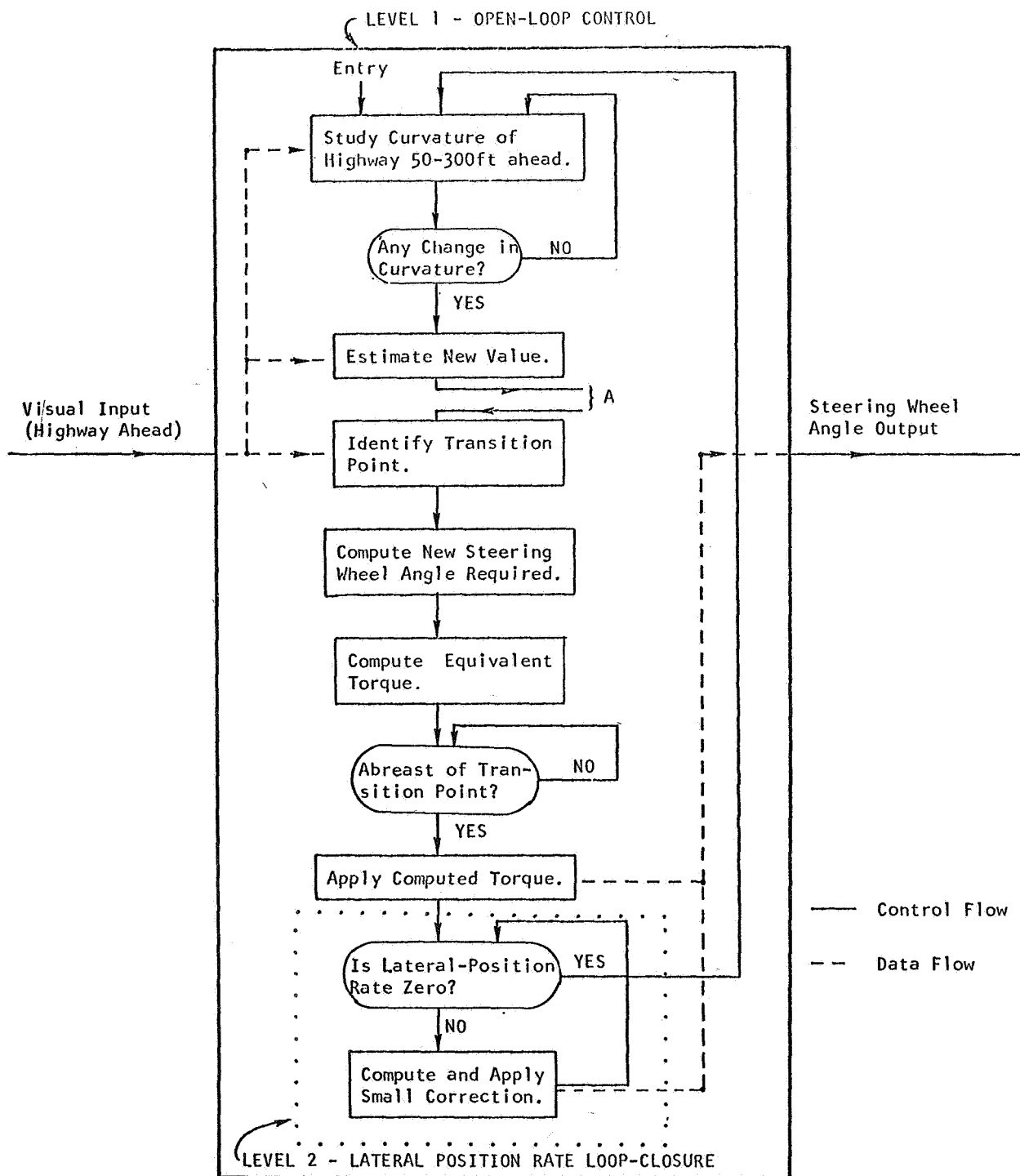


Figure 5. —Flow chart for open-loop preview control of path curvature in highway steering. (A speed-control subroutine may be interpolated at point A.)

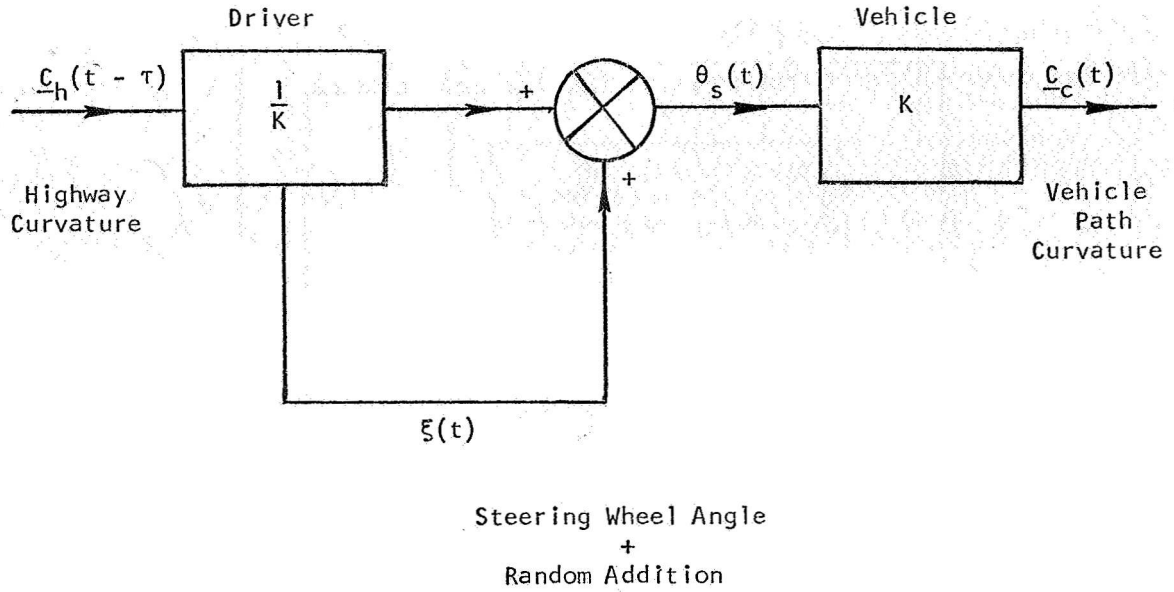


Figure 6.—Equivalent flow diagram for open-loop preview control of path curvature routine shown in figure 5.

With an idealized vehicle as described on the previous page this yields the fundamental equation of vehicle steering

$$\frac{d^2 y_c}{dt^2} = v_c^2 \cdot K \cdot \theta_s \quad (8)$$

The steering task calls for the car's lateral position y_c to be kept equal to a desired or reference lateral position y_r , normally generated internally by the driver. Direct continuous negative feedback utilizing an estimate of y_e of course results in an unstable system, since the idealized open-loop response is

$$\frac{y_e}{\theta_s} = \frac{v_c^2 K}{s^2} \quad (9)$$

In the nonideal case there is also a tandem second-order lag above 1.0 hertz, making the position worse.

However, given the highly nonlinear cost-function associated with y_e , the reduction in low-frequency error power achieved by the open-loop curvature-control policy and the moderate high-frequency power normally present in the two forcing functions, there appears to be no need for a tight loop operating directly on y_e . We postulate a tight loop (high repetition rate or sampling frequency) operating on lateral position rate or on one of the intermediate variables which are by geometry proportional to it. These are relative path angle $\alpha_{rel} = \alpha_h - \alpha_c$ and the relative heading $\psi_{rel} = \alpha_h - \psi_c$ where ψ_c equals instantaneous angular heading of the car to a first approximation $\alpha_c = \psi_c$. At low curvatures and low lateral accelerations relative headings ψ_{rel} can presumably be estimated with good precision (perhaps 3 bits) by direct comparison of reference points or lines on the vehicle with distant points on the

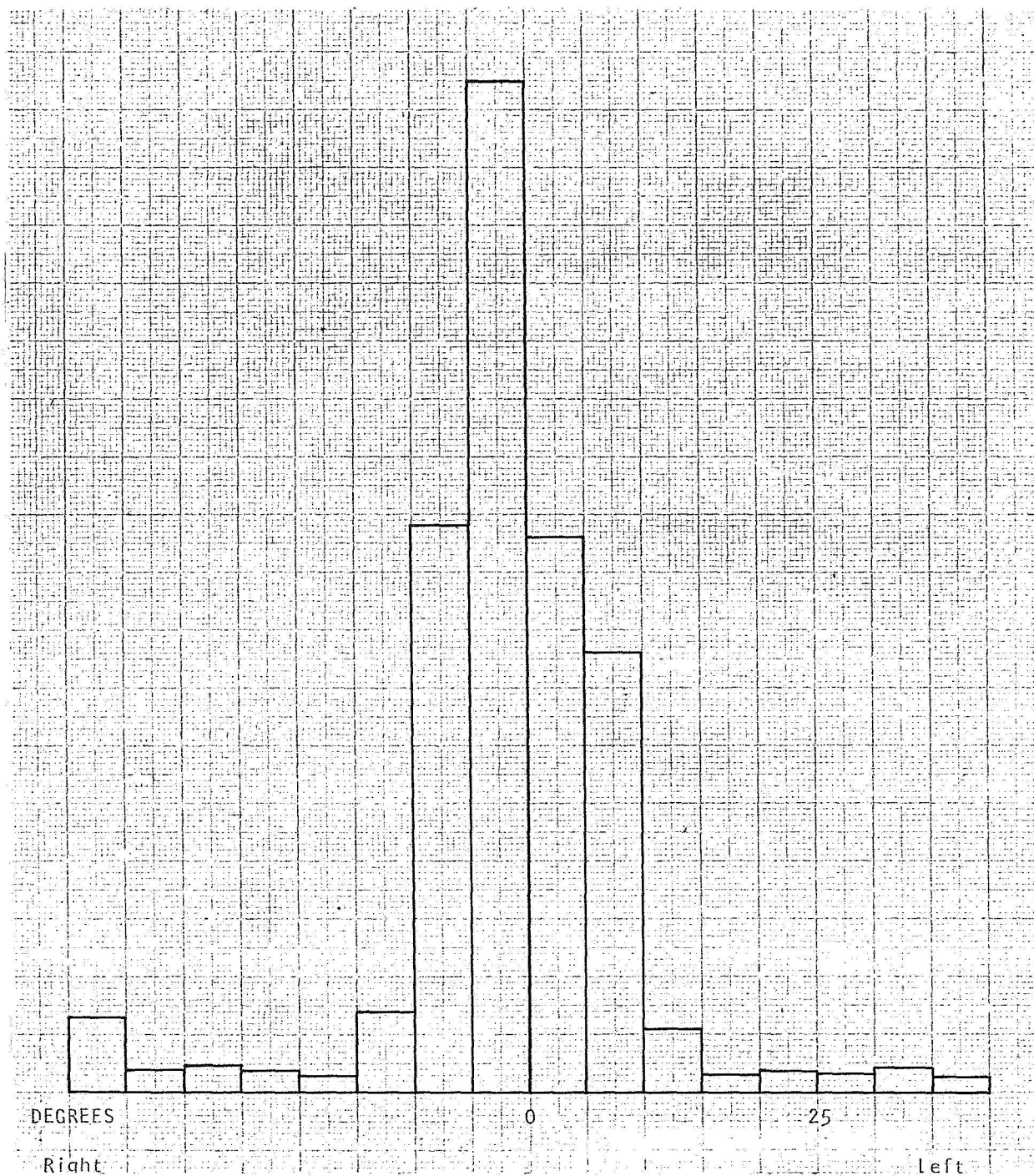


Figure 7.—Horizontal distribution of eye fixations in steering.
A low-curvature country highway (Interstate 680 between
Diablo and Walnut Creek, Contra Costa County).

highway or in the surrounding scenery. At high lateral accelerations vehicle sideslip complicates this picture, and ψ_{rel} ceases to be proportional to dy_c/dt . The angle α_{rel} itself would then be better estimated directly. This can be done by using the "center of expansion" of the visual field as described by Gibson and Crooks (ref. 14) and Biggs (ref. 15).

However the estimate is obtained, the level 2 loop would be closed by computing and applying negatively proportional increments to the steering wheel angle initially set by the open-loop curvature policy. A sampling frequency on the order of 1 to 2 increments per second would normally suffice.

The functional importance of relative heading estimation has been underlined by our experience with a fixed-based simulator, in which we have found it essential to provide directional cues in the visual field. In their absence subjects are simply unable to maintain normal control, while in their presence most subjects will proceed to "drive" in a normal manner. We have no direct data on the use of center-of-expansion as distinct from relative-heading inputs.

The above policy breaks down on curves, where relative heading can no longer be directly estimated due to lack of highway-related directional cues. Center-of-expansion judgments are also disturbed by the lateral (angular) velocity of all points in the visual field due to the vehicle's rotation. While it is possible that the latter cue might yield enough data to implement the policy described above at a lower level of accuracy (higher threshold), a third more plausible alternative exists in lateral-position rate, which can still be estimated from "shear" velocity at the boundary between hood and pavement in the visual field. This is highly correlated with relative path angle and might be expected to replace it when the latter fails. If this actually happens higher lateral-position error would result on curves than on straightaways due to the probably lower precision of first derivative as compared with direct estimates of variables. Whichever of these three possible second-level policies is feasible at any particular time, the very simple flow-chart of figure 8(a) would suffice to implement it. The corresponding block diagram is shown in figure 8(b).

Since the driver devotes attention to other than steering tasks, this particular loop will probably not be cycling continuously, but may be entered after variable time spent in other processing. This will have the effect of reducing the equivalent low-pass filter break frequency, so that we would expect the closed-loop performance to show time-variable amounts of high-frequency error power, depending on what other tasks are being time shared with steering (refs. 16 and 17).

We have obtained evidence for a level 2 loop of this type in a test-car steering experiment using a quasi-random forcing function injected between the driver's output and the actual steering wheel. This was implemented by means of our "double steering wheel" technique that simulates random side-thrust applied to the car. The lateral-position error was recorded by means of a TV camera error sensor. Preliminary results show that the driver introduces slightly less than the theoretical 6 dB/octave increase in gain below about 0.5 hertz, with a flat response above this frequency (fig. 9). This result is consistent with a control policy based on relative path angle or its equivalent, together with a driver-produced first-order filter with break frequency about 0.5 hertz. This is not inconsistent with a 1-second sampling rate, but we have no direct evidence either of the occurrence of sampling or of its frequency. However, further analysis of the data should reveal details of the driver's output and hence permit inferences as to his sampling activities.

LEVEL 3: CLOSED-LOOP CONTROL OF LATERAL-POSITION. —Since the precision of estimation of relative heading, relative path-angle or lateral-position rate are all limited to 3 bits per observation and will thus have a definite threshold, the second-level control

policy described above will fail to precisely null out relative-heading errors, though it will considerably reduce their amplitude and hence restrict the first derivative of lateral position to low levels, in other words act as a low-pass filter. Given initially zero lateral-position error y_e , only minor infrequent corrective action will suffice to maintain lateral position at its reference value y_r .

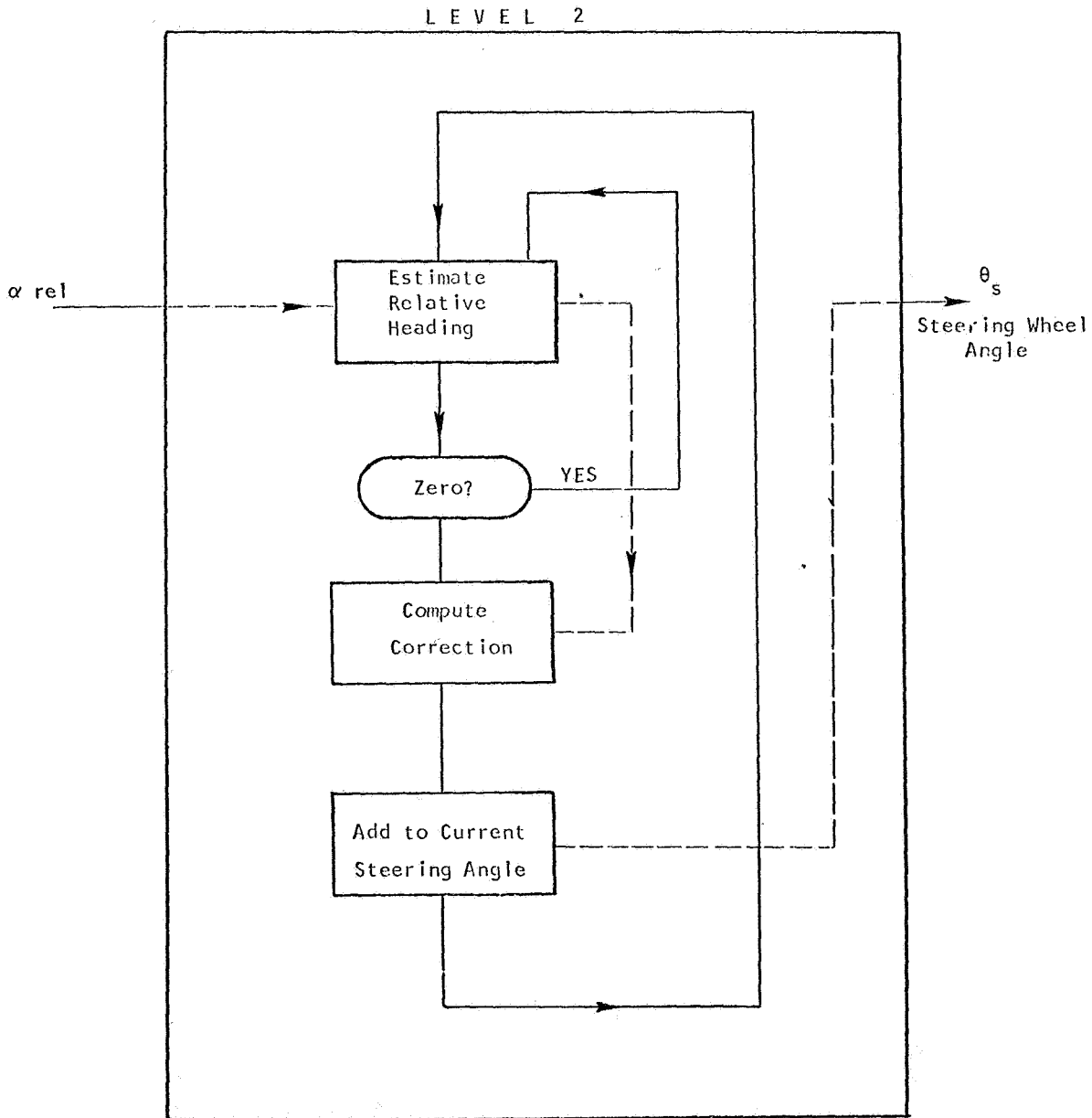


Figure 8(a). — Flow chart for level 2—Closed-loop compensatory regulation of relative heading.

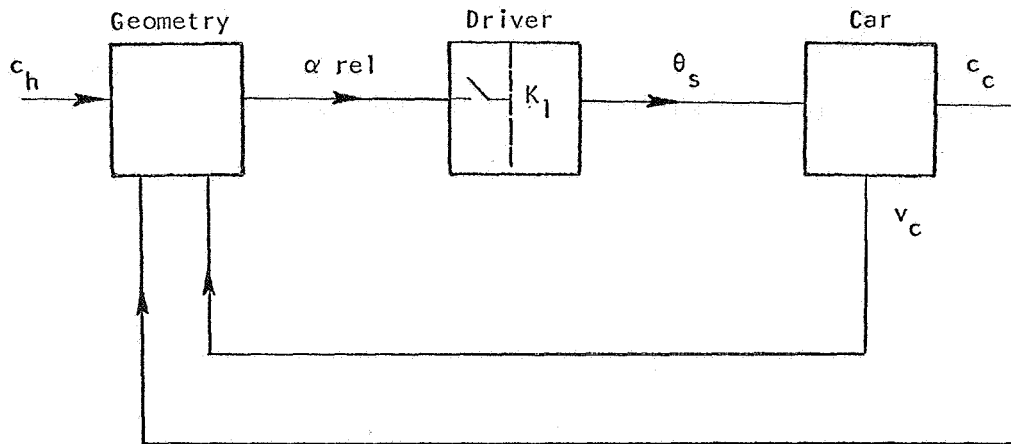


Figure 8(b). —Block diagram corresponding to figure 8(a).

A third-level policy to achieve this calls for direct estimation of lateral position y_c and its closed-loop adjustment to whatever reference value y_r is desired by the driver. As shown previously, this necessarily creates a second-order or higher order control system, but the low amplitude and low frequency of the residual forcing function applied to it, together with a relatively high error tolerance means that feedback gain can be kept low, for a low frequency of transient oscillation; a low sampling-rate will also suffice. Nevertheless we expect steady low-level oscillation to occur. This analysis is consistent with our observation of a low frequency periodicity in the 0.05 to 0.1 hertz region both in records of straight-highway simulator driving, and in actual studies of steering behavior on freeways.

To summarize the foregoing discussion, a model of the driver's steering behavior based on known parameters and constraints has been proposed viable for all normal highways and traffic conditions. It assumes a three-level hierarchy of control policies that reduce the difference between desired and actual path of the car, the former being defined jointly by the topography of the highway and the driver's preference for lateral positioning.

(1) Curvature of the highway ahead is estimated and the vehicle's path-curvature adjusted to conform to it by proportional change of steering wheel angle. This assumes that the driver "knows" the steady-state gain parameter relating steering wheel angle to path curvature.

(2) Residual error in car heading (or path angle) relative to the highway is nulled out by a negative proportional feedback with medium break frequency creating a first-order heading-control system and stabilizing the first derivative of lateral position.

(3) Lateral-position drift due to residual (subthreshold) discrepancy between highway direction and car path angle is nulled out by proportional feedback with low break frequency, creating a second-order lateral-position control system. This produces transient low-amplitude oscillation at frequencies on the order of 0.05 hertz.

It will be noted that no use has been made of lateral-acceleration data available from the driver's vestibular system and from the "seat of his pants." This is because in a well-behaved car on good surfaces lateral acceleration is dependent jointly on steering wheel angle and forward speed, both known quantities. Hence under these circumstances the vestibular information is redundant and could only serve to verify the occurrence of vehicle motions already commanded by the driver via the steering wheel. However in skids and slides the car's motion will no longer be exactly predictable from a knowledge of steering-wheel angle, so that vestibular data will presumably become useful in such cases. It will

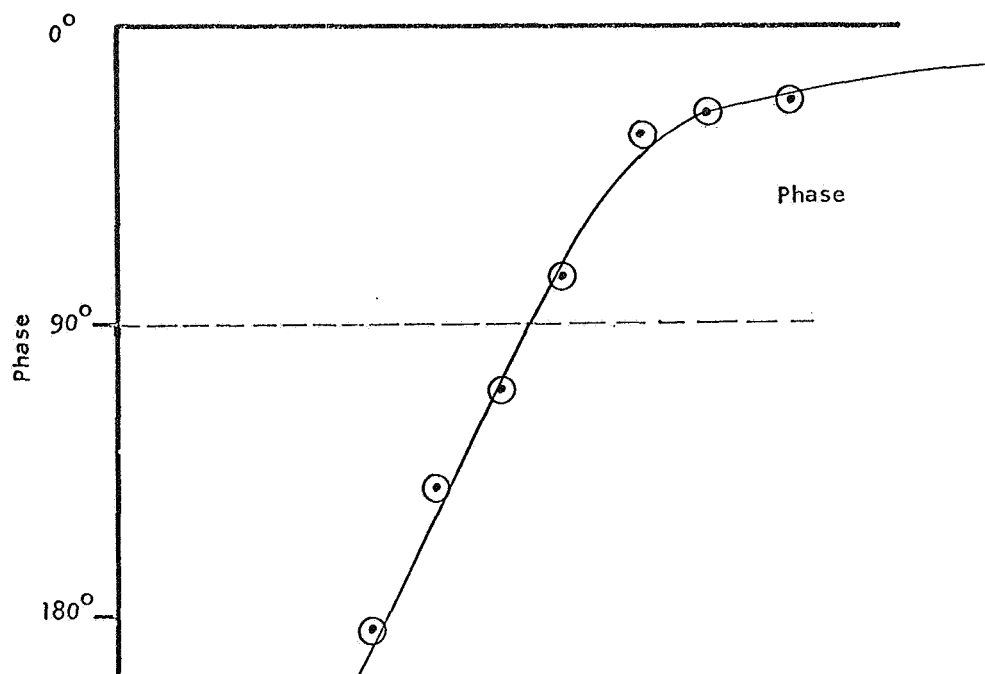
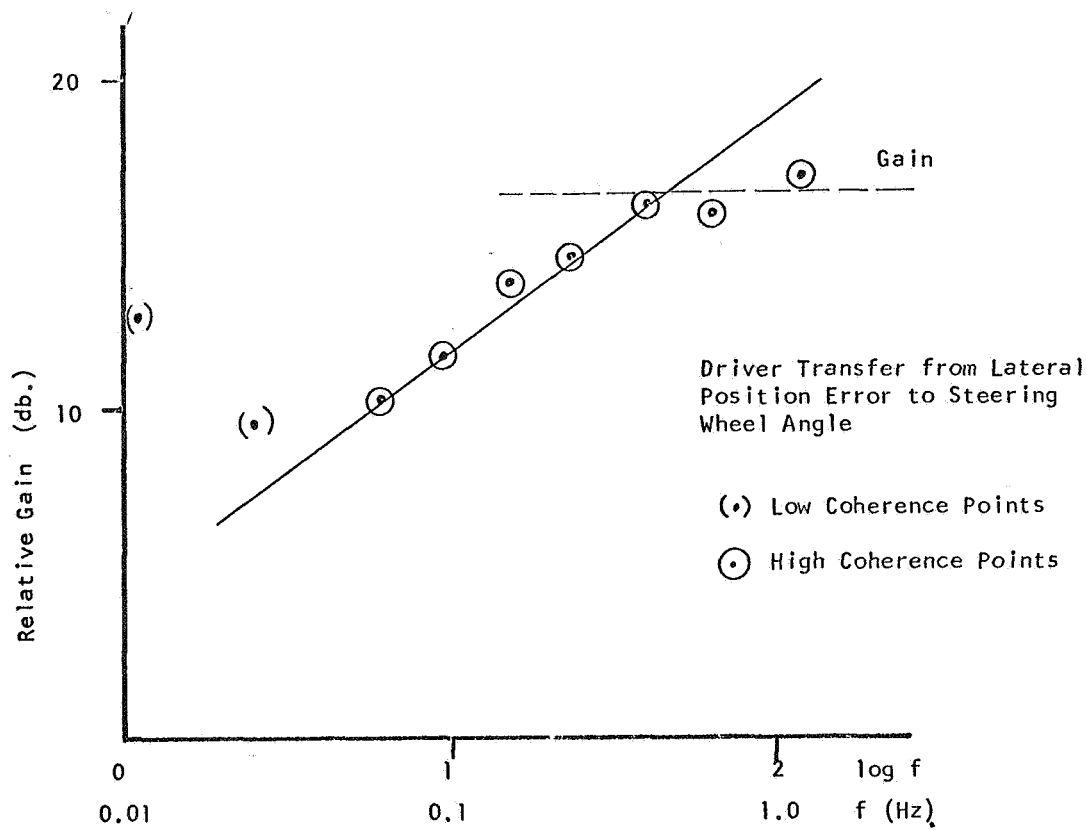


Figure 9.—Results of one subject in straight ahead driving with quasi-random forcing function introduced using the double steering wheel. 40 mph. S.T.I. forcing function.

also be noted that except for level 2 operating on curves all these policies call for proportional response, and none of them call for good high-frequency response. Thus the driver's response lag appears to be unimportant. A key question appears to be what are the actual and feasible minimum repetition rates (or sampling frequencies) at each level. While this certainly depends on the exact forcing function being applied at a particular time and on the various thresholds, it seems that level 1 and level 2 could operate with about a 1-1 ratio and a scan rate around 1 per second while level 3 would require perhaps 1/5 the scan rate. Data of Senders et al. are relevant here (ref. 18).

A suitable flowchart to implement all the levels together is shown in figure 10(a), and the corresponding sampled-data control system in figure 10(b). No claim is made that this model will fit all drivers all the time, but we believe it to be a realistic basis for further research aimed at estimating parameters and detailing driver response to different environments. We also feel that it may provide a useful framework for analyzing skill acquisition, since the levels could naturally be implemented in reverse order (3, 2, 1) with progressive improvement in stability, precision, and high-frequency response. At extremely high skill levels such as exhibited by racing drivers we would expect the precision of level 1 control to be such that relatively little error remains for the lower levels to remove. This might perhaps indicate the need for a level 0, using purely preprogrammed steering outputs whose timing alone would be adjusted to match the highway curvature.

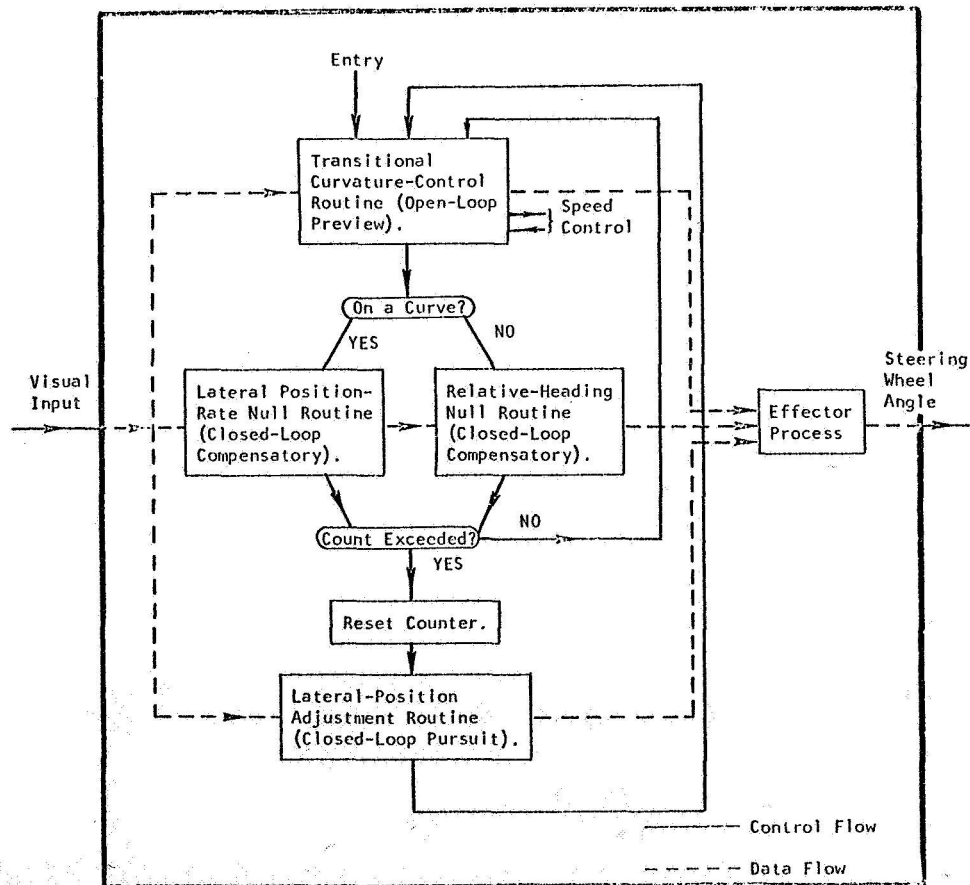


Figure 10(a).—Overall flow chart for driver information processing in vehicle steering.

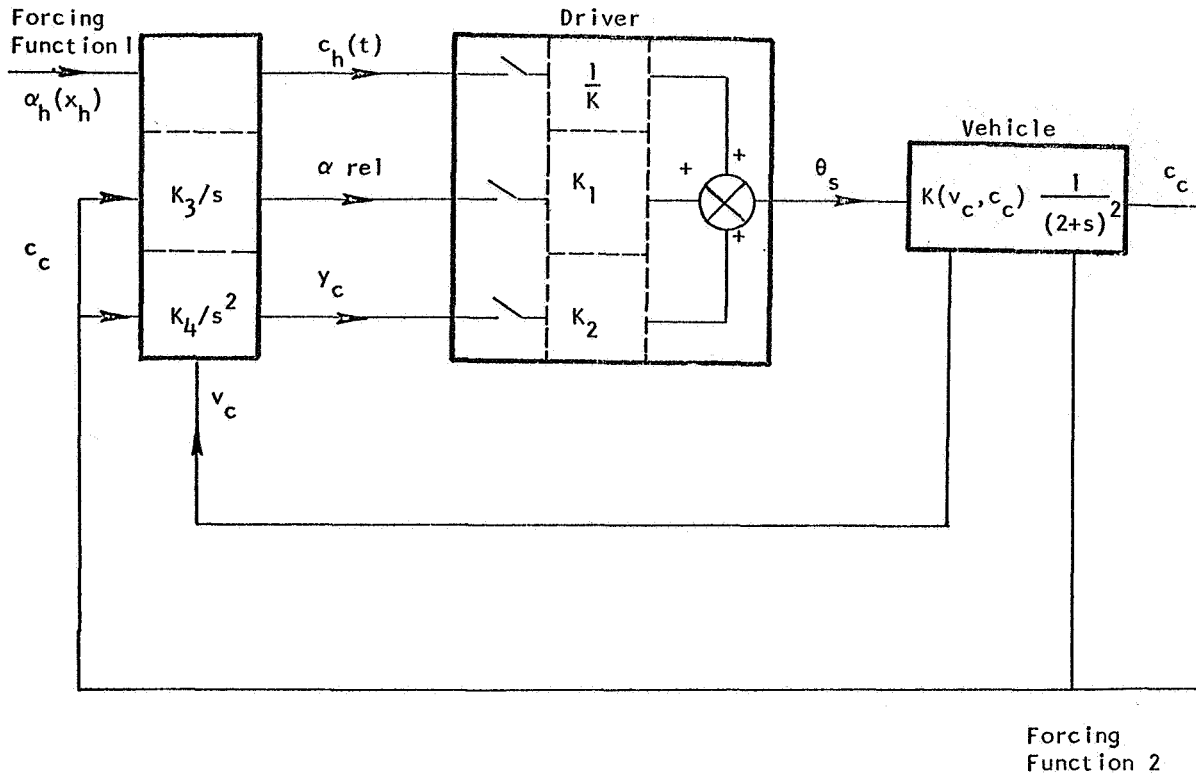


Figure 10(b).—Block-diagram equivalent of multilevel control policy shown in figure 10(a).

SPEED CONTROL IN RELATION TO STEERING

While this paper is primarily concerned with steering, the model presented above provides a natural framework for modeling the interaction between steering and speed control necessitated by the fact that steering fails on curves for which lateral acceleration a_r exceeds some threshold value $a_{r\max}$. Since speed control is implemented with long vehicle lags (ref. 10), the constraint on v_c implied by the condition

$$a_r = c_c \cdot v_c^2 \leq a_{r\max} \quad (10)$$

can only be met by feedforward control.

Since our level 1 already calls for estimates of highway curvature c_h , and since it is known that the vehicle's path curvature c_c cannot differ much from c_h , the necessary instantaneous $v_{c\max}$ can be readily computed by solving the equation

$$v_{c\max} \leq (1/c_h)^{1/2} \quad (11)$$

While it is unrealistic to suppose that the driver solves inequality (11) on-line, he may readily use values of $v_{c\max}$ "tabulated" against highway curvatures c_h , since only at most about 8 values of c_h can be discriminated. Clearly it is always possible to apply a safety factor by reducing the computed $v_{c\max}$ by any desired amount. A suitable subroutine to implement this requirement is shown in figure 11 and may be inserted at point A in figures 5 and 10(a).

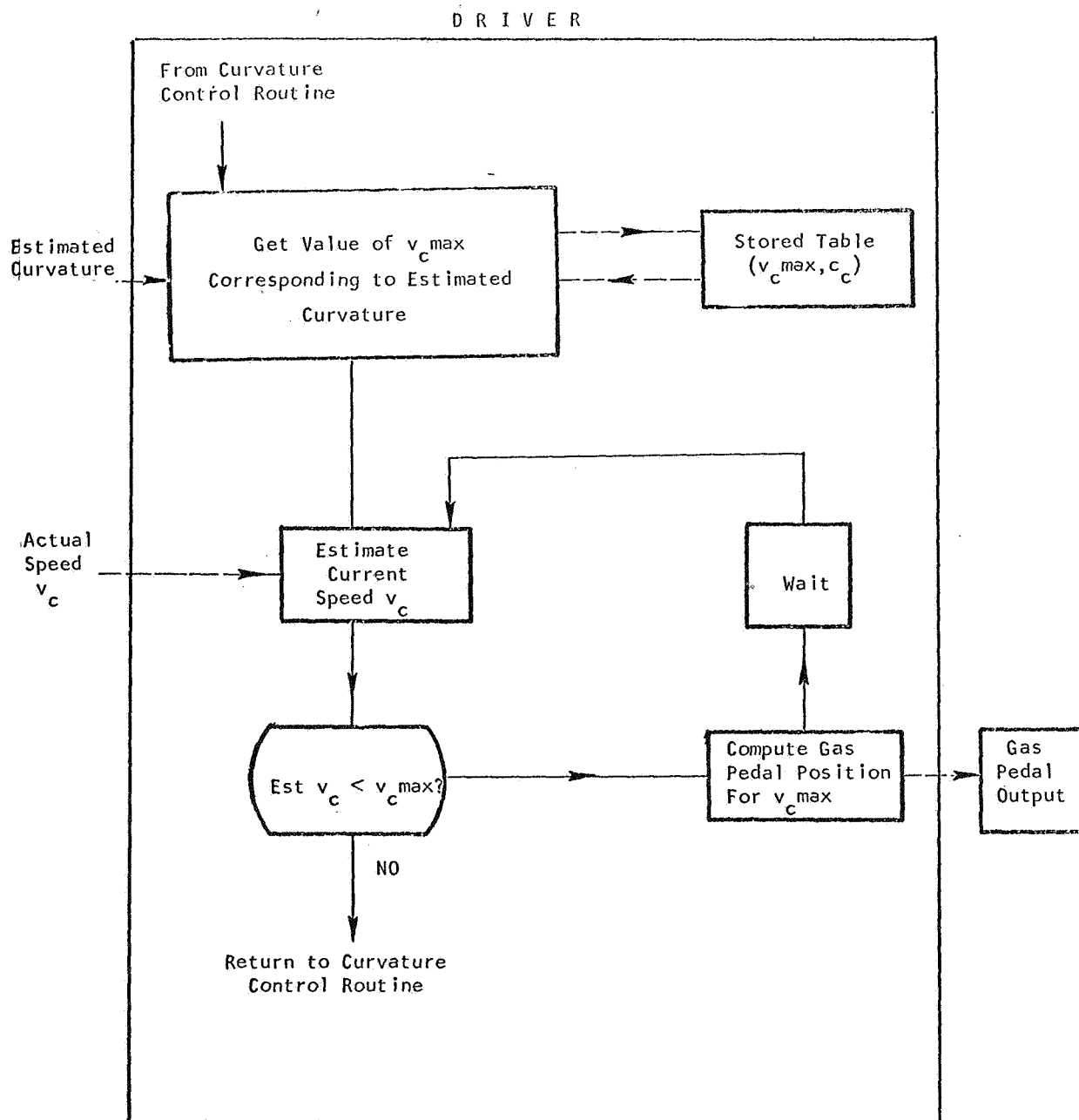


Figure 11.—Flow chart for speed-control adjustment routine to meet maximum lateral acceleration constraint.

REFERENCES

1. Rashevsky, N.: Mathematical Biophysics. Third ed., Dover, 1959.
2. Wohl, J.G.: Man-Machine Steering Dynamics. Human Factors, vol. 3, 1961, pp. 222-228.

3. Weir, David H.: Closed-Loop Directional Control of Automobiles. Paper presented at I. E. E. E. 8th Annual Symposium on Human Factors in Electronics. Palo Alto; May 1967.
4. Wierwille, W.; Gagne, A.; and Knight, J. R.: An Experimental Study of Human Operator Models and Closed-Loop Analysis Methods for High-Speed Automobile Driving. I. E. E. E. Transactions on Human Factors in Electronics, HFE8. vol. 3, 1967, pp. 187-201.
5. Miller, G. A.: The Magical Number Seven Plus or Minus Two; Some Limits on Our Capacity for Processing Information. Psychol. Rev., vol. 63, 1956, pp. 81-97.
6. Meiry, Jacob L.: The Vestibular System and Human Dynamic Space Orientation. Man-Vehicle Control Laboratory, M.I.T., June 1965.
7. Terao, M.: Quantization and Sampling Selection for Efficient Direct Digital Control. Instrumentation Technology, August 1967, pp. 49-55.
8. Segel, Leonard: Theoretical Prediction and Experimental Substantiation of the Response of the Automobile to Steering Control. Proceedings of the Automobile Division Institution of Mechanical Engineers. Paper no. 7, 1956.
9. Nordeen, D. L.: Vehicle Handling: Its Dependence Upon Vehicle Parameters. Soc. Automotive Eng. Preprint no. 5405, Feb. 1964.
10. Weir, D. H.; Shortwell, C. P.; and Johnson, W. A.: Dynamics of the Automobile Related to Driver Control, Technical Rept. no. 157-1, Systems Technology, Inc., July 1966.
11. Anon.: Motoring Which? Published by Consumer's Association, London, 1962-1968.
12. Szostak, H.: Investigation of the Dynamic Steering Response of a 1965 Ford Sedan Test Vehicle. Rept. HFT 67-2. Human Factors Laboratory, Univ. of California, Berkeley, 1967.
13. Pollack, I.: Private communication quoted by G. A. Miller, ref. 4., 1956.
14. Gibson, J. J.; and Crooks, L. E.: A Theoretical Field Analysis of Automobile Driving. Amer. J. Psychol., vol. 57, 1938, pp. 453-471.
15. Biggs, N. L.: Directional Guidance of Motor Vehicles—A Preliminary Survey and Analysis. Ergonomics, vol. 9, May 1966, pp. 193-202.
16. Cliff, R. C.: Interaction Between a Visual-Manual Compensatory Tracking Task and an Auditory-Verbal Shadowing Task. Rept. HFT 66-12. Human Factors Laboratory, Univ. of California, Berkeley, 1966.
17. Crossman, E. R. F. W.: Information Processes in Human Skill. British Medical Bulletin, vol. 20, 1964, pp. 32-37.
18. Senders, J. W.; Kristofferson, A. B.; Levison, W.; Dietrich, C. W.; and Ward, J. L.: An Investigation of Automobile Driver Information Processing. Bolt Beranek and Newman Inc., Cambridge, Mass., Rept. 1335, April 27, 1966.

III. HUMAN PERFORMANCE THEORY

11. Human Information Processing Rates During Certain Multiaxis Tracking Tasks With a Concurrent Auditory Task

*T. E. Wempe and D. L. Baty
Ames Research Center, NASA*

A series of experiments were conducted to determine the information processing rates of several subjects performing one- and two-axis compensatory tracking tasks with a secondary auditory task. The experimental variables were the order of controlled element dynamics, the forcing function, and the addition of a secondary task.

Human information processing rates decreased on each tracking channel with the addition of the second tracking channel or the secondary auditory task. Other than this effect, the information processing channels were additive like parallel channels until a limit in the total information processing rate was reached. This limit was related to the order of the controlled element.

INTRODUCTION

This study is a continuation of research (refs. 1 and 2) in investigating the utility of measures of transinformation (information processing rate in bits/sec) in describing and predicting human performance in tasks related to aerospace missions. It would appear that if future research is to span the gap, in the sense of predicting human limitations, between the typical laboratory experiment involving one or perhaps two tracking channels or some other set of relatively simple tasks and the multiplicity of dissimilar tasks occupying a pilot's attention, then some measure, such as transinformation, will be the medium. Certainly any hope of being able to describe workload for a complex set of heterogeneous tasks will be dependent on the validation of some generalized unit of measure such as bits/sec.

What is known about human transinformation capability in tracking tasks? Figure 1, a possible model of a human one-channel information processing system, is offered for discussion (some of the terms are borrowed from Melton, ref. 3). To an observer, the only evidence that information was processed appears in a comparison of the signal source and the system output. The information processed, as measured across the serial portion of this block diagram, cannot be any greater than the least information processed through any single block of the series. Conversely, each block in a series must have an information processing rate of at least the magnitude of the transinformation indicated between the system input and the system output.

Two prior studies (refs. 1 and 4) of single-axis tracking with a displacement controlled element have established the general shape of transinformation curves. When plotted against signal bandwidth, the curves were unimodal, having peaks of about 4 to 7 bits/sec

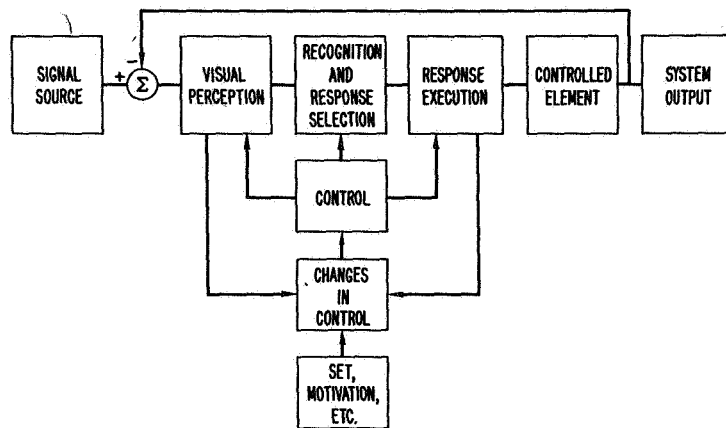


Figure 1. —Information-processing model of single-channel compensatory tracking.

(depending on subjects and forcing function shapes and whether the task was pursuit or compensatory) at a bandwidth of about 0.7 hertz, and were skewed toward the high bandwidth end of the scale. Additional evidence in another study (ref. 5) indicated that two-axis tracking with homogeneous dynamics (K/S^2) and either homogeneous (3.5, 2.5, or 1.5 rad/s) or heterogeneous forcing function bandwidths (3.5 and 1.5 rad/s) yielded essentially the same transinformation per channel as single-axis tracking (a slight loss of 0.1 bit/sec in the dual-axis task was attributed to "visual-motor interference effects").

The foregoing experimental evidence suggests that most of the human transinformation limitations in tracking are due to restrictions in visual perception and response execution (neuromuscular output). It is not until experimental data for heterogeneous dynamics (K and K/S^2 both at 3.5 rad/s) of that same report (ref. 5) are admitted that there is a suggestion that the box labeled "recognition and response selection" in figure 1 places a restriction on system transinformation. In this case a loss of 1 bit/sec in the K tracking axis could be attributed to interference from the large displayed error on the K/S^2 axis; however, the loss of 0.5 bit/sec in the K/S^2 axis suggests interference at a higher level mental process other than visual perception or motor response.

The aims of the present study were to verify the essentially parallel dual-tracking channel effects noted by Levison and Elkind (ref. 5) for both homogeneous and heterogeneous inputs when the controlled elements were similar and to replicate the effect of heterogeneous controlled elements. Also, a secondary auditory task, intended to separate perception and motor channels, was added to bypass the effects of visual-motor channel interaction.

SYMBOLS

K	gain of controlled element
S	Laplace operator used in defining controlled element
W_{eff}	effective bandwidth of forcing function

$$\left(\left[\int_0^\infty \phi_{ii}(f) df \right]^2 \right) / \left(\int_0^\infty \left[\phi_{ii}(f) \right]^2 df \right), \text{ Hz}$$

$S(f)$	signal power at frequency f
$N(f)$	noise power at frequency f
ω_n	natural frequency of the filter used to generate the forcing functions
Φ_{oo}	output power spectral density
Φ_{ii}	input power spectral density
Φ_{io}	input to output cross-power spectral density
λ	performance measure of the critical task device ($\approx 1/\text{effective reaction time}$)
(S)	single-axis tracking only
(S+)	single-axis tracking with concurrent secondary auditory task
(D)	dual-axis tracking only
(D+)	dual-axis tracking with concurrent secondary auditory task
Rel (E-N)	relative error with noise component removed

$$\text{Rel N} \quad \text{relative noise} \left(\sqrt{\frac{(\text{mean square error}) - (\text{mean square noise})}{\text{mean square input}}} \right)$$

$$\left(\sqrt{\frac{\text{mean square noise}}{\text{mean square input}}} \right)$$

TESTS AND PROCEDURES

TEST SETUP.—Both manual control tasks, the primary task of visual compensatory tracking and the secondary task of pressing the audio button, were performed by the subjects while seated inside a small portable cab. The doors were closed during the runs and every effort was made to eliminate disturbing outside noises. The small ventilating fan in the cab provided sufficient background noise to mask any conversation in the room. A light inside the cab was just intense enough to light the frame of the scope. The subject sat erect with his forehead on a headrest that maintained his eyes approximately 50 cm from the face of the scope.

Continuous Compensatory Tracking Task.—The elements for this task were displayed on a 14-inch Dumont 436 oscilloscope. A 1/4-inch reference circle remained centered on the scope and a 3/8-inch cross hair follower could be electronically driven anywhere on the face of the scope.

The task forcing functions were provided by a multichannel FM magnetic tape system (Ampex CP-100). The filtered output of a low-frequency gaussian noise generator had been prerecorded on magnetic tape. The specifications of the noise generator were: gaussian amplitude distribution within ± 1 percent and output spectrum uniform to ± 0.1 decibel from 0 to 35 hertz. The recorded signal was then shaped by a second-order filter having the transfer function $G/(S^2 + 1.4\omega_n S + \omega_n^2)$, providing a -80 dB/decade power spectrum beyond the break frequency for a forcing function. The filter gains at each bandwidth were adjusted so that the standard deviations of the recorded signals were essentially the same. The

effective bandwidths¹ used for this study were 0.12, 0.47, and 1.88 hertz which correspond to ω_n settings of 0.5, 2.0, and 8.0 radians/sec. The rms scope deflection of the uncorrected input was slightly under 1 cm, which at 50 cm from the eye gave an uncorrected visual angle of approximately 1° rms.

Error control was provided through compatible movements of a two-axis MSI model 438 sidearm controller. A specially made flexible control stick 9.0 cm long with a 1.5-cm-diameter ball on the end was used in place of the standard rigid one. The stick was mounted upright and would deflect 1 cm at the tip with a 6×10^5 dyne side force. The controller provided two independent electrical outputs, one proportional to the horizontal and the other proportional to the vertical component of deflection. The stick was allowed to move freely in both axes in all parts of the experiment. When the task was either vertical only or horizontal only, the error indication was clamped electronically at zero displacement in the inactive channel (one exception to this will be noted later in this report).

Three controlled element dynamics were used for this task: displacement (1.6×10^{-5} cm error displacement per dyne of stick force), velocity (7.8×10^{-5} cm/sec error displacement per dyne), and acceleration (7.8×10^{-5} cm/sec² error displacement per dyne).

Discrete Auditory Secondary Task.—Two audio tones (350 and 600 hertz) were presented in random order through the cab speaker. A small box with two push buttons was strapped to the subject's left leg. His task was to press the button that corresponded to the audio tone being presented, using his index or middle finger of his left hand. When he pressed the correct button, the tone ceased. One tone per second was the rate of presentation. The two buttons for this task were mounted at one edge of a $5 \times 13 \times 15$ cm metal box, which was large enough for the subject to rest the heel of his left hand while pushing the buttons. The switches required approximately 1×10^5 dyne of force to close contacts.

TEST SUBJECTS.—Four male college students served as subjects for this set of experiments. They were all right handed, had normal corrected vision, and had not participated in any prior tracking experiments. Subject C held a current pilot's license.

One of the experimenters who had fairly close contact with the subjects during the experiments described them as follows:

Subject A appeared to be a fairly easy-going type who did not seem to get very involved with the experiment.

Subject B was always interested in how well he had done and indicated displeasure when he thought he had done poorly. He was the only subject who stated that he had set some error criteria as a goal. At the end of experiment I, he left the project for personal reasons.

Subject C never appeared to have more than a mild interest in performing the tasks and occasionally demonstrated undesirable independence by not following instructions exactly.

Subject D seemed to be the slowest to comprehend the instructions and the slowest to attain stable performance at the tasks; however, he always seemed interested in improving his performance.

PROCEDURE

Instructions.—For the compensatory tracking task, the test subjects were told "Your score on this task is directly related to how close you can keep the cross to the circle throughout the entire run." For the audio task, they were instructed to respond to each tone

¹Effective bandwidth is defined as the bandwidth of a rectangular power spectral density that has the same area and variance as the power spectral density being described.

within 1 second or less of the onset of the tone. They were told that their score was the number of correct responses minus the number of incorrect responses divided by the total number of tones presented during the test period. They were told that they were not scored on how quickly they responded, as long as they responded within the 1 second interval.

When the tracking task and the auditory task were presented together, the subjects were not told how to weigh the two tasks. They were only told to do their best on both.

The subjects were informed of their prior day's performance at the beginning of each day and were urged each time to better their scores.

Performance measures.—Two scoring procedures were used for the compensatory tracking task. An online relative RMS error score was computed for each run to give a day-to-day indication of subject progress and for informing the subjects of this progress. The other procedure was to digitize directly and store on magnetic tape the system input, output, and error signals for each axis being tracked. These data were used in the offline computation of transinformation measures.

For the auditory task, the number of input signals, number of correct responses, and the number of incorrect responses were recorded during each run. The method of scoring from these values has been stated in the instructions to the subjects.

Training and Experimental Design.—The general experimental plan was to have each subject train with the same controlled element in one and two axes, with and without the secondary auditory task. Then experiments I and II were conducted. Each subject was then trained with a different controlled element, and experiment III was conducted. The final part of the plan, experiment IV, was to measure performance with the two learned controlled elements in two axes combined. Table 1 summarizes the details of the experiments. Table 2 presents the details of the Latin squares used in experiments I and IV (experiment III was similar to experiment I and is not shown).

Generally, two subjects were run per day, one resting in the subjects' waiting room while the other was being tested, so that there was always at least 1/2 hour between each of the three daily sessions for a given subject. The runs were 3-1/2 minutes long when W_{eff} of the inputs were 0.12 or 0.47 hertz, and 2-1/2 minutes when the input W_{eff} was 1.88 hertz. During a session, the rest periods were 1-1/2 minutes between runs. A seventh run that was not recorded was added to each session of the last day of the initial training period and maintained throughout experiments I and III. The condition for the seventh run was randomly chosen from the preceding six conditions (see table 2) of each session.

For experiment IV each session contained six runs, each condition being run twice in succession (see table 2). The results of the second run of each pair were recorded.

Data Reduction.—The input and output signals for each of the tracking tasks were digitized online (sampled from track-and-store units at the rates of 10/sec for $W_{\text{eff}}=0.12$ and 0.47 hertz and 20/sec for $W_{\text{eff}}=1.88$ hertz). For each pair of input and output signals, 1800 samples per channel were obtained for each run and stored on magnetic tape for offline computation. Cross correlation and autocorrelation values with 90 lags and subsequent power spectral densities were computed. The transinformation values were obtained by the following formula:

$$\begin{aligned} \text{Transinformation} &= \int_0^{\infty} \log_2 \left[1 + \frac{S(f)}{N(f)} \right] df \\ &\approx \Delta f \sum_f \log_2 \left[1 + \frac{S(f)}{N(f)} \right] \end{aligned}$$

TABLE 1. --EXPERIMENTAL PLAN

Phase	Subject			
	A	B	C	D
Initial training period	$\frac{K}{S^2}$	$\frac{K}{S^2}$	$\frac{K}{S}$	K
	72 runs (6 days) at $W_{eff} = 0.12$ Hz, single and dual axis, with and without secondary task Three days with same conditions as Exp. I	Same as A	54 runs (3 days) at $W_{eff} = 0.47$ Hz, single and dual axis, with and without secondary task Three days with same conditions as Exp. I	90 runs (5 days) at $W_{eff} = 0.47$ Hz, single and dual axis, with and without secondary task Three days with same conditions as Exp. I
Experiment I	$\frac{K}{S^2}$	$\frac{K}{S^2}$	$\frac{K}{S}$	K
	Three days, three sessions per day, seven randomly sequenced runs per session (See table II)	Same as A	See A	See A
Experiment II	$\frac{K}{S^2}$		$\frac{K}{S}$	K
	5 runs. Dual-axis tracking, single axis input, Horiz. and Vert. at $W_{eff} = 0.12, 0.47$ and 1.88 Hz. (No secondary task)	Terminated	See A	See A
Second training period	$\frac{K}{S}$	---	K	$\frac{K}{S^2}$
	Three days with same conditions as Exp. III		Two days with same conditions as Exp. III	Four days with same conditions as Exp. III
Experiment III	$\frac{K}{S}$	---	K	$\frac{K}{S^2}$
	Same schedule as Exp. I (Similar Latin Sq. Design)		Same schedule as Exp. I	Same schedule as Exp. I

TABLE 1. -- CONCLUDED

Phase	Subject			
	A	B	C	D
Experiment IV	$\frac{K}{S^2} \text{ \& } \frac{K}{S}$		$\frac{K}{S} \text{ \& } K$	$K \text{ \& } \frac{K}{S^2}$
	All dual-axis tracking (No secondary task) 54 runs (3 days), heterogeneous, W_{eff} or controlled element, or both. (See table II)	---	See A	See A

TABLE 2. -- LATIN SQUARE DETAILS

Experiment I

	Subject A (K/S^2)			Subject B (K/S^2)			Subject C (K/S)			Subject D (K)		
	1	2	3	1	2	3	1	2	3	1	2	3
S	0.12	0.47	1.88	0.47	0.12	1.88	1.88	0.47	0.12	1.88	0.12	0.47
E	.47	1.88	.12	1.88	.47	.12	.12	1.88	.47	.12	.47	1.88
S	1.88	.12	.47	.12	1.88	.47	.47	.12	1.88	.47	1.88	.12
I												
O												
N												

Note: Each recorded session consisted of six runs of the following conditions presented in random order: (S) vertical axis, (S) horizontal axis, (S+) vertical axis, (S+) horizontal axis, (D) and (D+).

Experiment III was similar to Experiment I with the subjects having different controlled elements and a different randomized Latin square.

S E S S H O N

Note: Lower case letters refer to the test conditions for each cell of the Latin squares above and are described in the cells for Day 1.

Note: Lower case letters refer to the test conditions for each cell of the Latin squares above and are described in the cells for Day 1.

where

$$1 + \frac{S(f)}{N(f)} = \frac{\Phi_{oo}(f)}{\Phi_{oo}(f) - \frac{|\Phi_{io}(f)|^2}{\Phi_{ii}(f)}}$$

RESULTS AND DISCUSSION

PRELIMINARY TESTS AND TRAINING DATA.—Though it was not intended that this paper should be concerned with individual differences, it was necessary to be cognizant of these differences where they would have a major effect on conclusions. Just prior to the sequence of experiments of this paper, each subject made 170 runs on the critical task device (ref. 6) which gives a measure that is inversely related to effective reaction time. Subsequent to the current experiments, each subject made an additional 120 runs on this device. The averages of the 40 runs just before, and the 40 runs just after the current experiments, are presented in figure 2 along with the single-axis performance averages of experiments I and III. The improvement in critical task performance shown in this figure was not significantly different from that noted between any other adjacent averages of 40 runs, indicating no transfer of training from the extensive tracking experience (a total of 20-1/2 hours of actual tracking for each subject completing the experiments) of the current sequence of experiments. The ranking of the subjects by use of the critical task device, which fulfills more of the criteria for eliciting maximum effort from the subjects than the current experimental setup (ref. 7), was not very different from the ranking by the data of experiments I and III. Only subject D changed position in the ranking. Differences, as will be discussed later, were partly due to motivational differences, particularly for the relatively easy single-axis task without the secondary task.

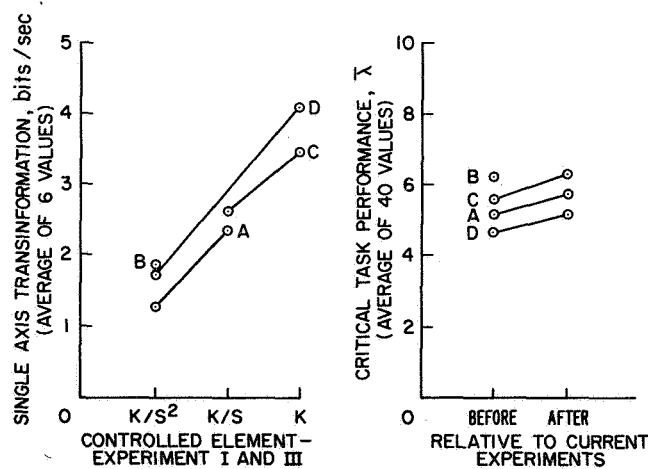


Figure 2. —Comparison of subjects' critical-task performance and performance in experiments I and III.

The training data from the first day were recorded and transinformation measures computed. These data are presented in figure 3 along with the comparable values from experiment I. Subject A had difficulty in learning the tasks that were more complex than the single-axis (K/S^2) condition. His results on experiment I look much like subject B's first efforts. Subject B established stable performance more quickly and showed rapid and consistent improvement in the most difficult condition of two-axis plus secondary task. Subject C (K/S) showed essentially equal performance in the single- and dual-axis tasks with only little improvement. He had distinct trouble when the secondary task was present and showed the most improvement in the single-axis with secondary task. Subject D (K) showed considerable improvement throughout the training which was consistent with his low rank on the critical task, his slowness to catch on, and the seriousness with which he approached each of the tasks. Like subject C, he initially showed about equal proficiency at the single- or dual-axis tasks, but had difficulty with the secondary task.

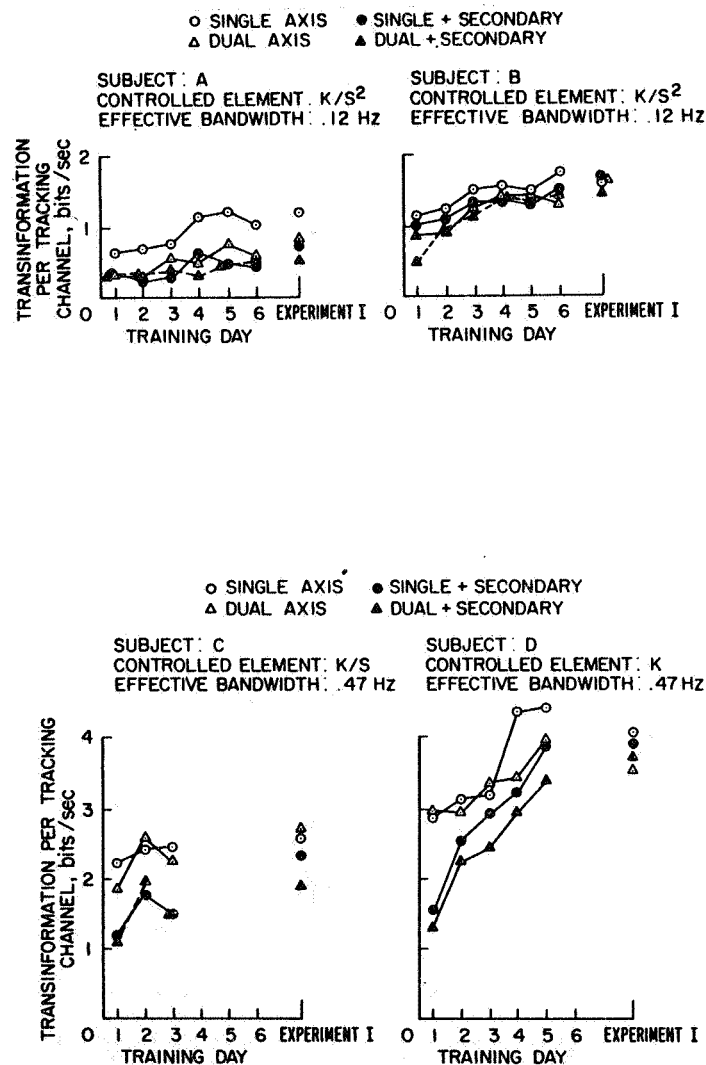


Figure 3. — Transinformation during early training and comparable data from experiment I.

HOMOGENEOUS TEST CONDITIONS

Secondary Auditory Task.—Performance for the secondary task was fairly uniform across all conditions and the few trends that were noted were attributed to subject differences and were not considered significant. The average for the secondary task when it accompanied the K/S² controlled-element tracking task was 0.88 bits/sec (though subject D's performance averaged 0.97 bits/sec with this controlled element); the average with the K/S controlled element was 0.95 bits/sec; and the average with the K controlled element was 0.97 bits/sec.

Days, Sessions, and Sequence Effects.—Experiments I and III were planned to counterbalance the effects of variations in performance from day to day, from session to session, and throughout the test sequence during a session. Although the averages of the controlled variables were not affected, there were some significant results from these unwanted variables. On experiment I, subject B's performance evidenced a significant (0.05) drop² of approximately 0.25 bit/sec on the third session of each day and a significant (0.01) drop on the last day; subject C showed a consistent decrease in performance (0.01) of 0.4 bit/sec from the first session to the last session; subject D displayed erratic behavior showing a session's effect (0.05), a sequence effect (0.05), a days by sequence interaction (0.05), and a sessions by sequence interaction (0.05). By the time the three subjects who continued the experiments reached experiment III, none of these effects was present.

Analysis of Transinformation Data.—The per channel transinformation results (averages of six data points or three data points when the horizontal and vertical channels were significantly different and are shown separately) of experiments I and III are shown in figure 4.

To study the effects of the secondary task, a table of differences ((S) minus (S+) and (D) minus (D+)) was prepared for each session. The purpose was to minimize the variability due to days and sessions mentioned before. (It was noted later that the results would have been the same had an analysis of variance been performed directly on the transinformation measures.) A similar table was prepared for (S) minus (D) and (S+) minus (D+) to investigate the effects of adding the second tracking channel. For both of these tables, the horizontal and vertical channels were combined to achieve some level of generality. An analysis of variance was performed, and the data were combined where the confounding of significant effects did not result. These results are presented in table 3.

With the exception of the data for subject D at the K controlled element task, the interactions indicated in table 3 for the K/S and the K controlled element seemed excessive. Some interaction was attributed to the subjects' reacting to a more lively task, such as subjects A and C performing better on the (D) task than on the (S) task (see fig. 4), but this did not seem to account for all of the interactions among the experimental conditions.

The significant interactions among these data could also have been caused by a ceiling on the total transinformation for each run. A ceiling would cause the (D+) transinformation values to be lower than would be expected if only the effect of the additional tracking channel and the effect of the added secondary task were influencing this measure.

Total Transinformation.—Figure 5, the total (sum of all active channels) transinformation for each run, shows evidence of ceilings in some cases. Both subjects C and A operating the K/S controlled element (figs. 5(b) and 5(c)) were unable to add the secondary task to the dual channel task to achieve a higher total transinformation than when they were performing at the dual task alone. Their apparent ceiling for the K/S controlled element was approximately 5-1/2 bits/sec. At the K controlled element task (fig. 5(d)), subject C performed noticeably poorer in total transinformation on the (D+) condition than on the (D)

²Levels of significance resulting from analysis-of-variance statistics are shown in parentheses.

condition suggesting some functional degradation in addition to reaching a ceiling. These examples are considered evidence that a fairly low ceiling for total transinformation exists for the (D+) condition, although it may have been higher if these subjects had been more highly motivated. Subject D, a harder working subject, performing with the same controlled element, K, as subject C (fig. 5(e)), showed no indication that he had reached a ceiling at 8-1/2 bits/sec. This subject's early learning total transinformation data are shown in figure 6 to demonstrate that even for him total transinformation was limited when he was operating at a lower level of proficiency. None of the three subjects using the K/S^2 controlled element (fig. 5(a)) showed any evidence of a total transinformation limit at their highest average of 3-1/2 bits/sec.

Effects of Adding a Second Tracking Axis.—To determine the effect of adding the second tracking axis, the use of the differences between the (S+) and (D+) measures was not considered valid where an apparent ceiling on total transinformation would limit the (D+) values.

Thus, the table 3 values showing a single- to dual-axis loss of 0.26 bit/sec (0.01) for the K/S^2 task and a loss of 0.27 bit/sec (0.05) for subject D on the K task are considered

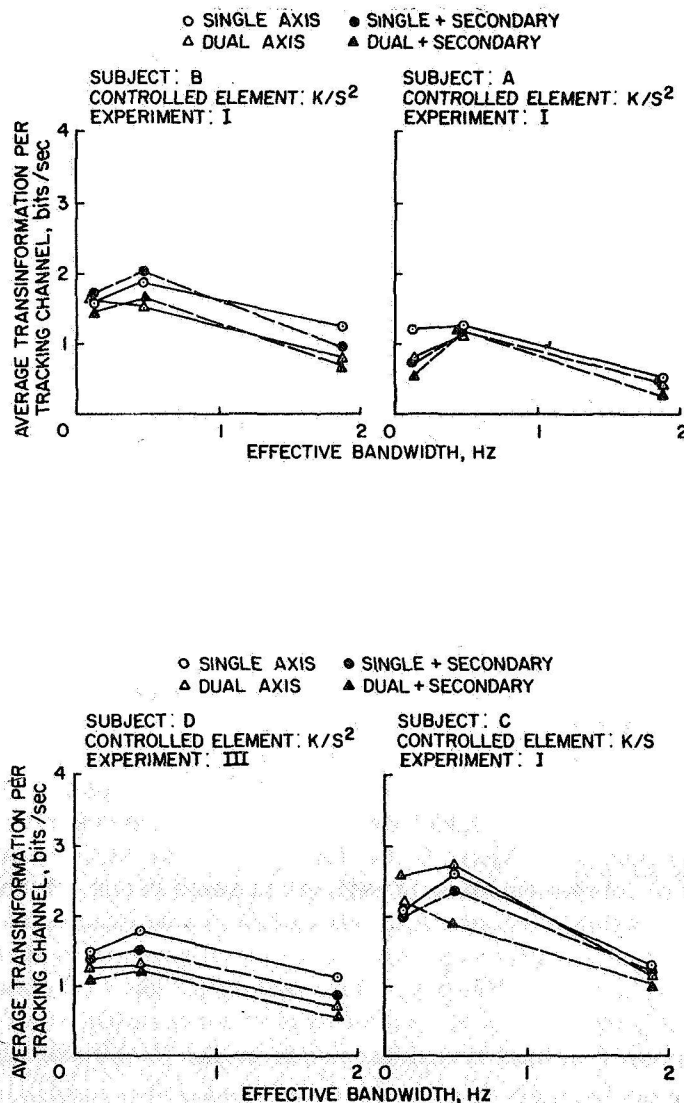


Figure 4.—Results of experiments I and III (averages of six values per data point, except where horizontal and vertical channels are separated).

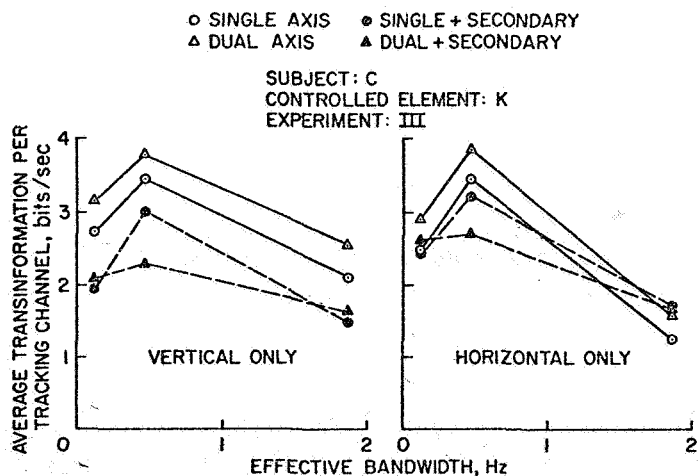
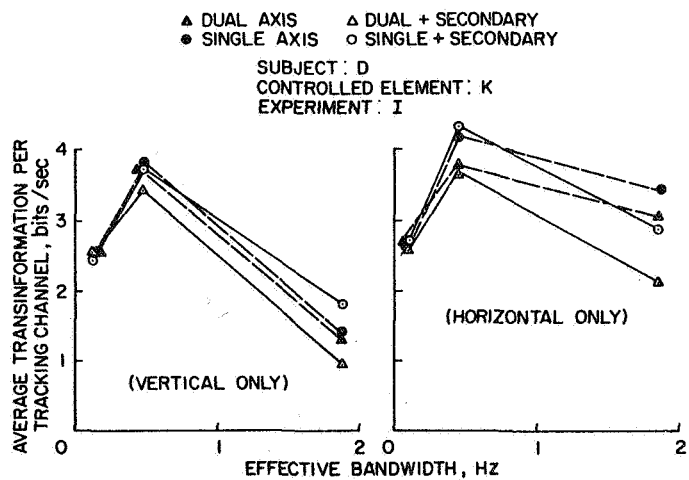
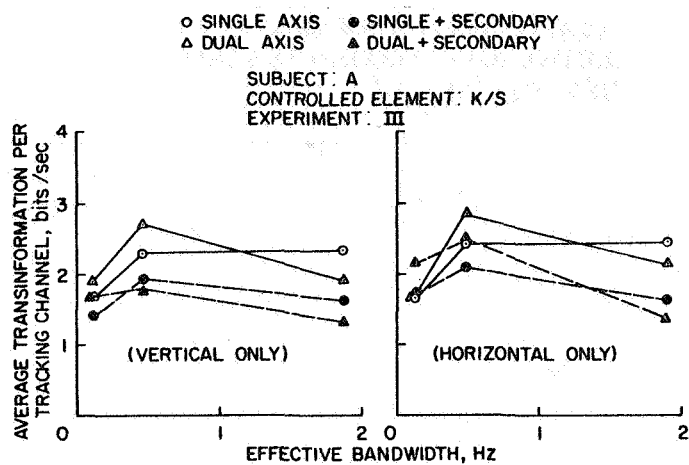


Figure 4. — Continued.

TABLE 3.—CHANGES IN TRANSFORMATION PER TRACKING CHANNEL
DUE TO THE ADDITION OF A SECOND TRACKING AXIS OR A
SECONDARY AUDITORY TASK

Controlled Element	Subjects	Condition	W_{eff} , Hz	Average per channel change, bits/sec
(a) Second axis added				
K/S^2	A,B,&D	(-) & (+)	All	-0.26**
K/S	A & C	(-) & (+)	0.12	.29*
			.47	.04
			1.88	-.23*
K	D	(-) & (+)	All	-.27*
	C	(-)	All	.57*
		(+)	All	-.14
(b) Secondary auditory task added				
K/S^2	A,B,&D	(S) & (D)	All	-.12**
K/S	A	(S) & (D)	.12	.03
			.47	-.49*
			1.88	-.72**
	C	(S)	All	-.13
		(D)	All	-.42*
K	D	(S) & (D)	All	.13
	C	(S)	All	-.29
		(D)	All	-.97**

(-) = no secondary auditory task
(+) = secondary auditory task present
(S) = single axis
(D) = dual axis

** = significant at 0.01 level

* = significant at 0.05 level

Overall average, second axis added = -0.11

Overall average, secondary auditory task added = -0.23

valid since there was no apparent ceiling for these results. The data for test conditions without the auditory task were used for the remaining comparisons. Subject C operating the K controlled element showed a gain of 0.57 bit/sec (0.05 from table 3), and subjects A and C operating the K/S controlled element showed no significant effect from the addition of the second axis.

From this it is concluded that, with the more demanding K/S² controlled element task, a loss of 0.26 bit/sec was noted when the second axis was added. (This conclusion is consistent with the findings of Levison and Elkind (ref. 5) who noted a significant single- to dual-axis loss of 0.1 bit/sec with a K/S² controlled element tracking task.) However, while the subjects were performing with the easier K/S or K controlled elements, this small effect was obscured by a tendency for the subjects to track even better when the second axis was added. That this small effect should have been present, even though not evident in the subjects' performance, will be demonstrated when the results of experiment II (two-axis tracking with an input in only one of the two tracking channels) are discussed.

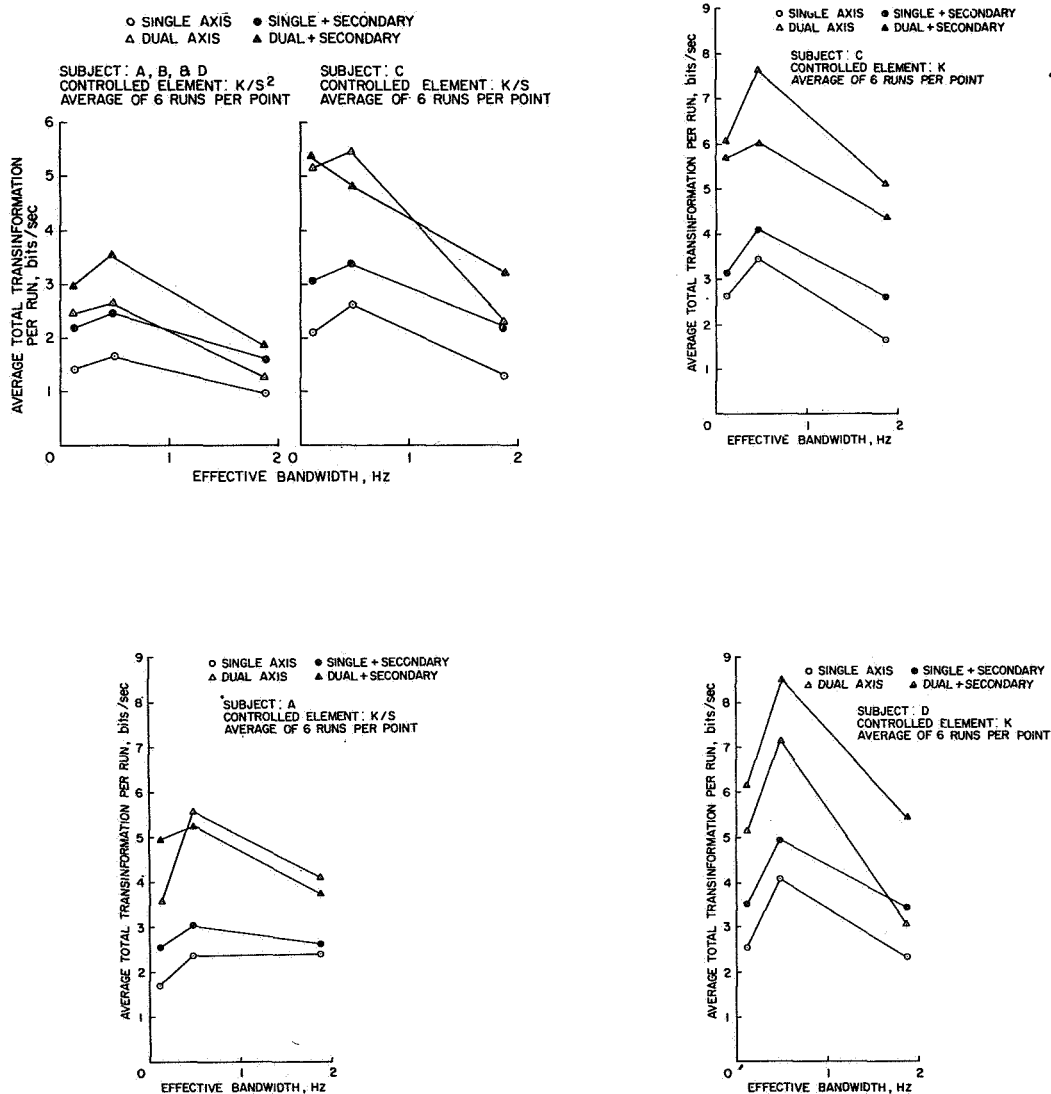


Figure 5. — Total transinformation per test condition— experiments I and III.

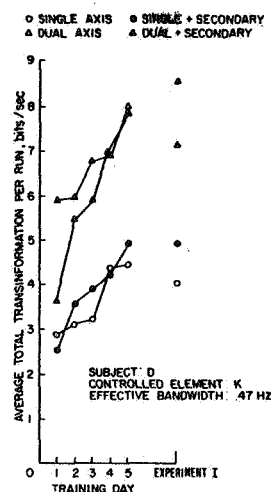


Figure 6.—Total transinformation during early training and experiment I.

Effects of Adding the Auditory Task.—When the effect of the addition of the auditory task was examined with the exclusion of dual-axis data where a ceiling on total transinformation was indicated, the data of table 3 for subjects A and C on the K/S and K controlled element task appeared uniform. An analysis of variance of the differences due to the addition of the auditory task to the single-axis tasks for all subjects and all controlled elements showed no significant effects from subjects, controlled elements, or bandwidths. The overall average loss in single-axis tracking from the addition of the auditory task was 0.18 bit/sec (0.01). Thus, it was concluded that the auditory task caused a small but significant loss.

Since the secondary task allowed a separate auditory perception input channel as compared to the tracking task shown in figure 1, and a separate response execution output channel (namely the other hand to activate the response microswitches), visual-motor interference was not likely to be the cause of this loss per tracking channel.

Figure 5, depicting total transinformation achieved for each experimental condition, verifies that the (S) to the (S+) loss was not due to a maximum transinformation limit in the recognition selection box of figure 1 since the (D) or the (D+) total transinformation measures were always higher than the (S+) measures. It was thought that sampling at the recognition and response selection function might account for this loss when the secondary task was added. This hypothesis was explored briefly by examining estimates of changes in the subjects' reaction times. The phase angles of the open-loop transfer function for the two subjects operating the K controlled element with the 0.47 hertz forcing function were measured at 1 hertz (where changes in reaction time would produce a relatively large change in phase angle, and changes in a lag time constant would produce a relatively small change in phase angle) and subjected to an analysis of variance.

The results showed that subjects C and D were somewhat different in their adaptations to the tasks for the K controlled element. Subject D showed no difference in apparent reaction time due to the addition of the auditory task, but he did show a small but significant (0.05) shift in average phase angle for the dual-axis task with or without the auditory task. This average phase angle shift from -121° to -126° would occur with an increase of 0.014 second in reaction time. The other subject, C, showed no change in average phase angle with the dual task, but he did show a consistent (0.01) shift from -127° to -136° when the auditory task was added to either the single- or dual-axis task. This difference would occur with an increase in reaction time of 0.024 second.

These observations seemed promising, particularly since subject D, who showed no reaction time effect from the addition of the secondary task with the K controlled element also showed no loss in tracking transinformation at that condition (see table 3), while subject C, who showed an apparent reaction time increase with the addition of the secondary task, at the same time showed a significant loss in tracking transinformation. Although these results are consistent with the notion that subject D was not sampling between the secondary task and the tracking task while subject C was sampling between these tasks, the early training data do not support this idea. When subject D's early training data were examined, his apparent reaction time still showed no change with the addition of the secondary task, but at that time he did show a definite loss in tracking performance when the secondary task was present (fig. 3). Thus, it appeared that sampling alone would not account for the entire loss in tracking transinformation when the secondary task was added.

Effects of the Controlled Element.—The difference in ceilings noted between the K and the K/S controlled element tasks (note that subject C showed different ceilings at each of these tasks, figures 5(b) and 5(c)) suggest that the total transinformation limit for tracking is related to the order of the controlled element, and that the recognition and response selection box of figure 1 has restrictions in total transinformation capability that are related to the need for acting on higher order derivatives of the input and output signals. Certainly there is a significant effect from the controlled element on the per channel transinformation of human trackers as emphasized in figure 7, with each change in the order of the controlled element producing a 1 bit/sec change in each tracking channel. This latter effect might be attributed solely to visual-motor noise resulting from the larger inherent error associated with higher order controlled elements; however, that would not explain the restriction in the total transinformation ceiling by the order of the controlled element. Further research is required to segregate and quantify these effects of the controlled element.

Effects of Forcing Function Bandwidth.—The effect of signal bandwidth on transinformation in tracking tasks has been noted before (refs. 1 and 4), and the current results agree with the data of those studies when the forcing function spectral shape, motivation of subjects, and display presentation are taken into account. With allowances made for visual-motor interference, it appears that tracking information processing channels can be added as though they were parallel, at least until the total information processing rate reaches some ceiling. The important point is that the box of figure 1 labeled "recognition and response selection," through which it is presumed that all information must pass, appears to have no limitation for the amount of transinformation required by a single tracking channel

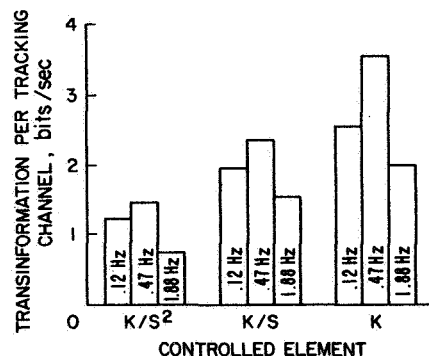


Figure 7.—The effect of controlled element on transinformation (averages of all subjects and all conditions from experiments I and III).

once a particular controlled element is specified, otherwise the channels would not appear to be parallel. Hence, it is not unreasonable that the principal restrictions in transinformation noted at low and high bandwidths for a given controlled element may be due entirely to limitations in the perception and response execution boxes of figure 1.

VISUAL-MOTOR INTERFERENCE.—When the plans for experiment II were established, it was intended that the results would apply toward explaining the (S) and (D) channel increase in remnant and related loss in per channel transinformation. The results of this brief experiment with two axes to control, but with an input in only one channel, are presented in table 4.

Though there appeared to be a linear relationship between the standard deviation of the no-input channel remnant and the standard deviation of the coherent error, it was not believed practicable to try to predict the increase in remnant due to the addition of a second axis and forcing function, particularly inasmuch as two of the subjects often showed a lower remnant in each of the dual-axis channels when the secondary task was not present. Table 4 was included to support the conclusion that there is visual-motor interference, and that a portion of the remnant in a second channel will be related to the coherent error in the first channel, at least with an integrated display and a single controller for two-axis control.

TABLE 4.—VISUAL-MOTOR INTERFERENCE IN TWO AXES TRACKING WITH AN INPUT IN ONLY ONE AXIS

Input W_{eff} , Hz	Controlled element					
	K/S ²		K/S		K	
	Axis with input Rel.(E-N)	Other axis Rel.N	Axis with input Rel.(E-N)	Other axis Rel.N	Axis with input Rel.(E-N)	Other axis Rel.N
0.12	0.28	0.47	0.14	0.07	0.12	0.06
.47	1.27	1.09	.57	.14	.41	.09
1.88	1.36	1.15	1.34	.39	.98	.13

Rel.(E-N) = RMS relative error with noise component removed.
Rel.N = RMS relative noise.

Note: There were two replications for each condition having forcing functions of 0.47 Hz and 1.88 Hz.

HETEROGENEOUS TEST CONDITIONS.—The purpose of experiment IV was to determine if the requirement for a different mode of responding to an additional tracking channel would place any special restrictions on the transinformation for each channel. Because display gain effects on transinformation are not understood, all the input signals (forcing functions) used in the experiments of this study were adjusted to have about the same standard deviation in amplitude. Thus, with different bandwidths or different controlled elements, the two-axis tracking error generally would not be equal in the horizontal and vertical axes. The channel with the largest error was considered to be the primary tracking channel, and the other, the secondary channel. Evidence for adjustment interaction would appear as a decrease in the transinformation of the primary channel, while it was anticipated that the secondary channel would be degraded by visual-motor interaction as well as adjustment interaction. If there were any effects at all from the heterogeneous conditions, it was expected that the least effect would occur with the heterogeneous bandwidths where only the values of the parameters of the assumed human transfer function need be different. It was expected that the most pronounced effect would accompany the heterogeneous controlled element conditions where the order of the response mode, as well as parameter values, would most likely be different.

Heterogeneous forcing function bandwidth.—Table 5 summarizes the results of the heterogeneous bandwidth data for each controlled element. The primary axis transinformation data were in almost every case higher than comparable experiment I or III dual-axis data, indicating that there was no degrading effect from the different adaptations required between the channels. Instead there apparently was an improvement in performance on the higher bandwidth channel, probably in an effort to equalize the displayed error in the horizontal and vertical directions. The secondary axis showed losses as expected.

Levison and Elkind (ref. 5) noted similar results from dual-axis tracking with heterogeneous bandwidths, although the bandwidths studied were over a smaller range; namely, 0.24, 0.40, and 0.56 hertz (rectangular spectra achieved by summing sine waves).

Heterogeneous Controlled Elements.—Table 6 presents the changes in transinformation for the heterogeneous controlled elements as compared to the dual-axis homogeneous controlled element condition (experiment I or III) for each effective bandwidth. The subjects apparently performed better than might be expected in the channel having the higher order controlled element and the larger displayed error. Thus, with a large difference in controlled elements that required different modes of response in each of the two axes, there was no degradation in transinformation in the primary channel. Poorer performance on the secondary channel (the lower order controlled element channel) as compared to the comparable experiment I or III dual-axis performance was anticipated because of the visual-motor interference effect described before.

It is concluded from the results of these heterogeneous bandwidth and controlled-element experiments that, at least for a low level of total transinformation, no measurable degradation in information processing rates was attributable to the subjects adopting different response modes for each axis of a two-axis integrated display and controller tracking task.

CONCLUSIONS

On the basis of the results of several experiments designed to study human information processing rates in manual control tasks of varying complexity, the following conclusions are indicated.

TABLE 5.—DIFFERENCE IN TRANSFORMATION BETWEEN HETEROGENEOUS INPUT BANDWIDTH TRACKING AND COMPARABLE HOMOGENEOUS DUAL AXES PERFORMANCE OF EXPERIMENTS I AND III

Subject	Controlled element	W_{eff} each axis, Hz		
		0.12	0.47	1.88
D	K/S^2	-0.45 bits/sec	0.48 bits/sec	
		-0.32		0.76
			-0.31	0.22
A	K/S	-0.24	-0.73	
		-0.27		0.31
			-0.61	0.86
C	K	-0.19	0.64	
		-0.83		-0.29
			-0.33	0.13

Average higher bandwidth channels = 0.26 bits/sec.

Average lower bandwidth channels = -0.39 bits/sec.

Note: There were three replications of each heterogeneous bandwidth task.

(1) Dual-axis tracking as compared to single-axis tracking with a K/S^2 controlled element evidenced a transinformation loss of 0.26 bit/sec in each channel. This loss was attributed to the visual-motor noise effect of one channel on the other. With the K/S or K controlled elements, evidence of this effect was not consistent among the subjects.

(2) The addition of an auditory secondary task, button pressing in response to one of two tones presented randomly at the rate of one per second, caused a transinformation degradation of about 0.2 bit/sec per tracking channel where a ceiling on total transinformation was not indicated. This effect was attributed to interference in a signal recognition and response selection function. Though there was some evidence for sampling at the recognition and response selection level, sampling alone did not seem to account for the effect.

(3) There was evidence that a ceiling on total transinformation, the sum of the transinformation of all active data processing channels, existed and that it was influenced by the order of the controlled element, with the lower order controlled element allowing a higher ceiling.

TABLE 6.—DIFFERENCE IN TRANSFORMATION BETWEEN HETEROGENEOUS CONTROLLED ELEMENT TRACKING AND COMPARABLE HOMOGENEOUS DUAL AXIS DATA OF EXPERIMENTS I AND III

Subject	W_{eff} , Hz	Controlled element		
		K/S ²	K/S	K
A	0.12	0.11	-0.71	
	.47	.58	-1.04	
	1.88	.20	-.70	
D	.12	-.45		-1.35
	.47	.65		-2.35
	1.88	1.00		-.20
C	.12		-.29	-.62
	.47		.54	-1.77
	1.88		1.17	-1.47

Average higher order controlled element = 0.39.

Average lower order controlled element = -1.13.

Note: There were three replications of each heterogeneous controlled element task.

(4) The order of the controlled element imposed a limit on the amount of transinformation for each channel. A loss of about 1 bit/sec was noted as the order increased from K to K/S and from K/S to K/S².

(5) When the dual-axis tracking task had either heterogeneous controlled elements or heterogeneous forcing function bandwidths, there was no evidence of a degradation in human transinformation performance that could not be accounted for by the visual-motor interference that one tracking axis had on the other.

REFERENCES

1. Wempe, T. E.; and Baty, D. L.: Usefulness of Transinformation as a Measure of Human Tracking Performance. NASA SP-128, 1966, pp. 111-129.
2. Baty, D. L.: Information-Processing Rate as Influenced by the Degree of Response Difficulty: A Discrete Tracking Task. NASA SP-144, 1967, pp. 157-164.
3. Melton, A. W.: Human Performance in Information Processing and Storage. Department of Psychology, Michigan University, Ann Arbor, Michigan.

4. Elkind, J. I.; and Sprague, L. T.: Transmission of Information in Simple Manual Control Systems. IRE Trans. on Human Factors in Electronics, vol. HFE-2, no. 1, March 1961, pp. 58-60.
5. Levison, W. H.; and Elkind, J. I.: Studies of Multi-Variable Manual Control Systems: Two Axis Compensatory Systems with Compatible Integrated Display and Control. NASA CR-554, 1966.
6. Jex, H. R.; McDonnell, J. D.; and Phatak, A. V.: A "Critical" Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time. Part I: Theory and Experiments With a First-Order Divergent Controlled Element. NASA CR-616, 1966.
7. Locke, E. A.; and Bryan, Judith F.: Goals and Intentions as Determinants of Performance Level, Task Choice, and Attitudes. AD 646 392, American Institute for Research, Silver Spring, Maryland, Feb. 1967.

12. Prediction and Decisionmaking in Manual Control*

*William R. Ferrell and Harry S. Cohen
Massachusetts Institute of Technology*

INTRODUCTION

A simple model of human decisionmaking can be applied to a common class of decision situations, a model sufficiently unrestrictive to have a good chance of representing actual behavior. Among the cases to which the model applies are those classed as signal detection, and a considerable body of experimental evidence substantiates the model's usefulness and its applicability to this and other psychophysical detection tasks. It appears possible to extend the model to include many examples of decisionmaking of practical importance in manual control, those in which an operator seeks to detect not a signal but his own potential success in carrying out an action.

THE MODEL

The decision model and the basic situation to which it applies can be set out very briefly.¹ The situation has the following properties:

- (1) There are two mutually exclusive states of nature, S and F, one of which must be the case.
- (2) There are observed data d , physical measures, relevant to whether S or F holds.
- (3) There are N alternatives, A, B, C . . . , one of which is to be chosen.
- (4) The data d generate a scalar p according to a stationary probabilistic process, which is independent of the N alternatives.
- (5) The entire p scale is partitioned into N mutually exclusive connected regions each of which is associated with one and only one decision alternative.
- (6) When a given observation d' is made, a scalar p' is determined as above and that alternative is chosen which is associated with the region into which p' falls. The p values separating the regions are termed cutoffs.

This simple structure can be related to human decisionmaking by (1) identifying p with the decisionmaker's subjective probability of S given the observation d , and (2) assuming

*This work was supported in part by the General Motors Corp. through the General Motor Grant for Highway Safety Research.

¹Although this statement is new, the model in various forms has been used by many writers. References to sources for the model as well as exposition of ROC curve analysis and its use with human signal detection may be found in D.M. Green, and J.A. Swets, *Signal Detection Theory and Psychophysics*, John Wiley & Sons, New York, 1966.

that there is a utility associated with each outcome, each combination of chosen alternative and state of nature. If the utilities are known, then by writing expressions for the expected utility of choosing each alternative using the subjective probabilities, cutoffs can be found that will maximize the subjective expected utility. Alternatively, any behavior following the pattern of the model can be interpreted as maximizing expected utility for some utility assignment.

THE RECEIVER OPERATING CHARACTERISTIC CURVE

The performance of a decisionmaker in a given set of circumstances is fully described by the frequencies of the various possible outcomes. When there are two states of nature, S and F, and only two alternatives, A (appropriate to S) and B (appropriate to F), then there are only three independent proportions. The probabilities $\Pr(A|S)$, $\Pr(A|F)$ and $\Pr(S)$ are generally used. When the first two are plotted one against the other, the model generates one point for each cutoff value on the scale, separating the regions associated with the two choice alternatives. The locus of points generated by shifting the cutoff is called the receiver operating characteristic (ROC).

The ROC is extensively used in the analysis of data from psychophysical experiments, especially the detection of signals in noise, for which theoretical curves defining the best possible performance can be obtained. The value of the ROC lies in the fact that it permits a separation of the observer's sensitivity to the true state of nature from the utility considerations affecting his choice. If the ROC curve can be characterized by one parameter then the detector can be given a single index of sensitivity. The operating point along the ROC reflects the decisionmaker's criterion for choosing one alternative over the other and is presumably determined by his utility for the various outcomes.

If the model described above holds, then there are two ways of generating the ROC curve, indirectly through an expression of p the subjective probabilities or directly by recording the outcomes of choices, the rewards for outcomes being varied from time to time to move the cutoff. In the indirect method, the subject observes the data and responds with a number r monotonic with p . For example, he might be asked to estimate his subjective probability of S in percent. From his response and knowledge of whether S or F was the case, the conditional distributions $\Pr(r|S)$ and $\Pr(r|F)$ are approximated. The experimenter then uses these to infer what would have been the result of decisions made according to the model with various cutoff values, as illustrated in figure 1.

The choice of the numbers to reflect p can, itself, be considered decisionmaking if one can assume that the utility of responding with a number r^{**} when S is the case is greater than that for responding r^* if and only if $r^{**} > r^*$. If this can be assured by external rewards, no assumption need be made about the direct accessibility of the quantity.

PREDICTING SUCCESS IN MANUAL CONTROL

In an auditory signal detection context an observer listens to a sound that may be either noise alone or noise containing a signal and, on the basis of what he hears, must make a choice between saying "noise" or saying "signal plus noise." The signal's presence or absence has been determined beforehand. The model has been found to be a very good description of human behavior for tasks of this kind. Manual control decisions often show a formal similarity to this paradigm. The operator observes the circumstances, the physical situation, and must decide whether to carry out a particular maneuver, one which he would certainly perform if he knew it would be successful. He is thus in the position of trying to detect from a variable set of physical circumstances the success of the action, even though

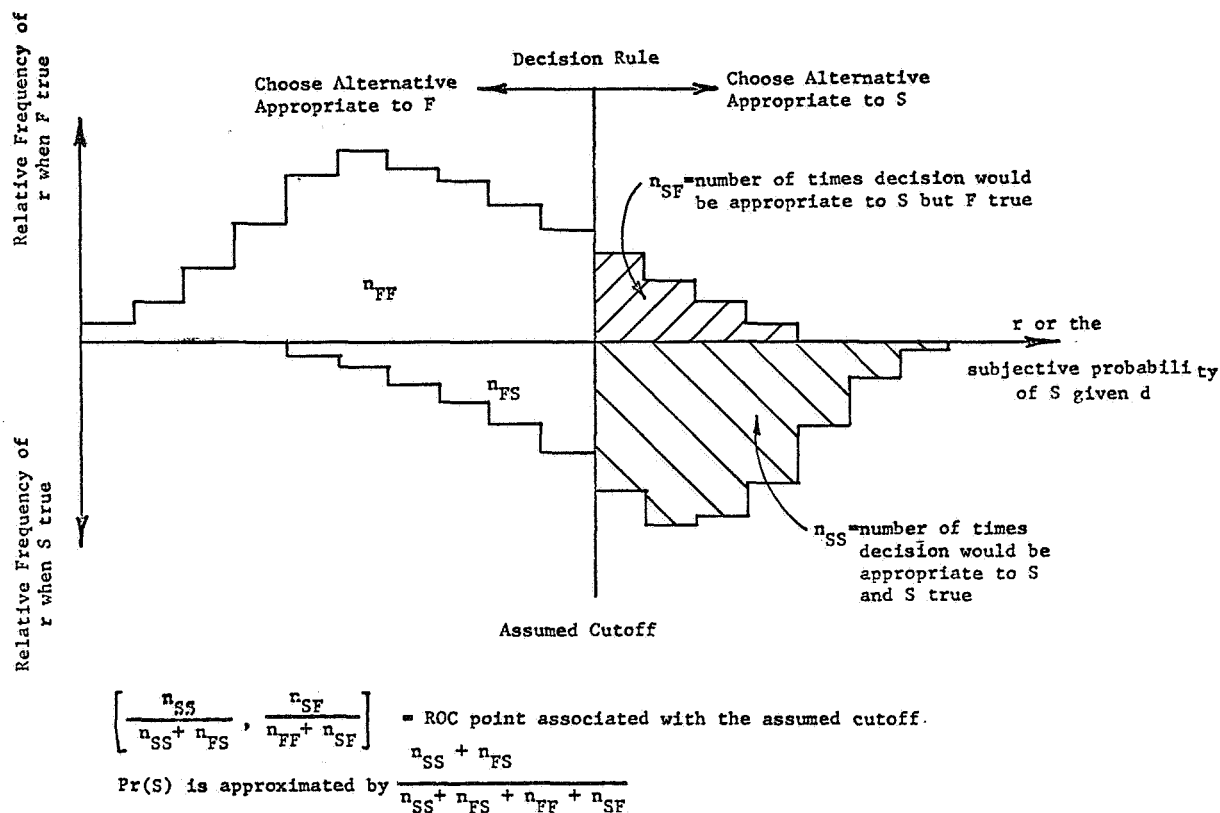


Figure 1.—Interpretation of data from probability estimation.

whether it will succeed has not yet been determined. Examples are whether to pass the car ahead, whether to continue a landing or make a go around, or whether to try to run an amber light.

In decision situations of this sort, three aspects of performance are relevant, (1) the decisionmaker's skill, (2) his ability to predict the result of his maneuvers, and (3) the relative importance to him of success and failure. If the given model holds, these components can be separately assessed for manual control decisions for which an ROC curve can be determined. Skill is represented by the probability of success, $\text{Pr}(S)$. The ability to predict is given by the ROC curve. The operator's effective utility for success relative to that for failure can be found from the operating point on the ROC and $\text{Pr}(S)$. If the decisionmaker operates according to the model with a cutoff p_0 , his expected objective utility is

$$\begin{aligned}
 & U_{SS} \text{Pr}(p \geq p_0 | S) + U_{SF} \text{Pr}(p \geq p_0 | F) [1 - \text{Pr}(S)] \\
 & + U_{FS} \text{Pr}(p < p_0 | S) \text{Pr}(S) + U_{FF} \text{Pr}(p < p_0 | F) [1 - \text{Pr}(S)]
 \end{aligned}$$

where p is his subjective probability and the utilities are defined by

	<u>S occurs</u>	<u>F occurs</u>
Response consistent with S	U_{SS}	U_{SF}
Response consistent with F	U_{FS}	U_{FF}

The quantities $\Pr(p \geq p_0 | S)$ and $\Pr(p \geq p_0 | F)$ are the ordinate and abscissa of the ROC. Rewriting the equation gives curves of constant utility as straight lines with slope equal to

$$\left[\frac{1 - \Pr(S)}{\Pr(S)} \right] \left[\frac{U_{FF} - U_{SF}}{U_{SS} - U_{FS}} \right]$$

If the subjective probability of S equals the objective probability of S, the second term is a measure of the utility for failure relative to that for success. It can be evaluated from $\Pr(S)$ and the tangent to the ROC at the operating point. Thus the three components of the decisionmaker's performance can be separately evaluated if the ROC analysis is valid for manual control decisions.

THE EXPERIMENTAL QUESTION

The usefulness of the ROC description of decision behavior in manual control tasks depends on whether real decisionmakers are adequately represented by the decision model. The model predicts that a decisionmaker has an ROC curve that is independent of the particular set of choice alternatives. If points on a decisionmaker's ROC graph generated by various sets of decision alternatives do not lie on the same ROC curve, his decision-making ability cannot be separated from the reward structure of the decision situation.

The existence of stable ROC curves for predicting success in manual control tasks was investigated by comparing points on an ROC graph generated by bet choices with points generated by probability estimation.

EXPERIMENT I

The experimental manual control task consisted of turning a knob to deflect an x-y recorder pen in the y direction past obstacles and into a target area as it moved at constant rate in the x direction. The dynamics were of the form k/S^2 . The obstacles and target varied randomly in both size and position. A typical configuration is shown in figure 2. The conditions were—

- (1) controller choosing bets
- (2) observer choosing bets
- (3) controller making probability estimates
- (4) observer making probability estimates

On each trial, subjects were first shown the particular obstacles-target configuration. In the bet choice sessions, subjects chose one of two bets, A or B, the payoff of which depended on whether the target was hit. On each trial one or another of the following payoff matrices was used:

	<u>Success</u>	<u>Failure</u>		<u>Success</u>	<u>Failure</u>
Bet A	9¢	0¢	Bet A	7¢	0¢
Bet B	5¢	1¢	Bet B	5¢	3¢

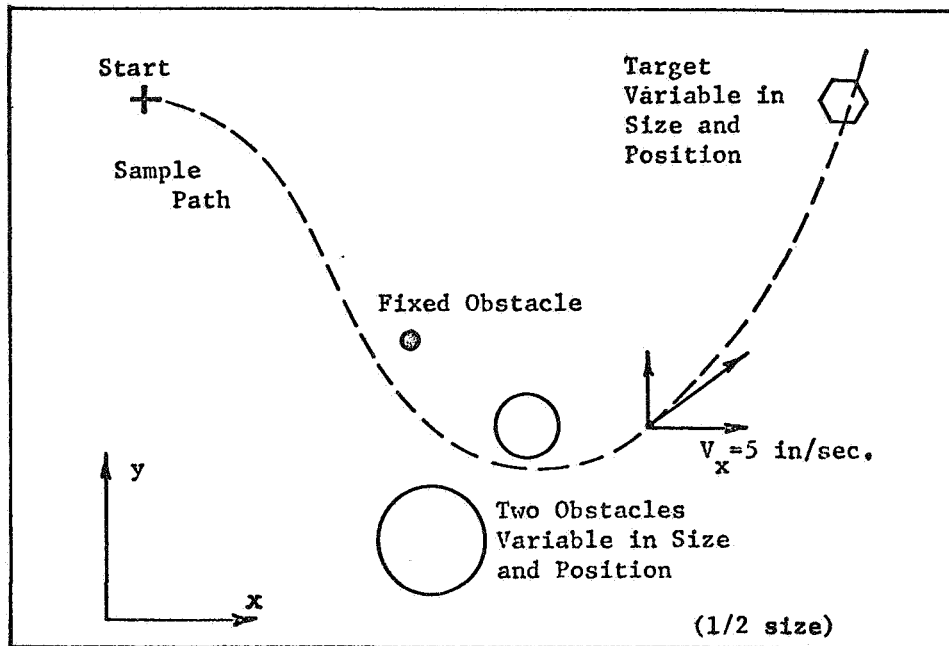


Figure 2.—Task layout for experiment I.

In no case did it pay to fail. In the probability estimation sessions, the subjects picked a number between 0 and 100 representing their feelings about whether or not the task would be performed successfully. Subjects knew that their pay would be determined by decisions chosen optimally using their estimates. Controllers and observers were paired. The controllers each performed the given task, after choosing or estimating, to determine the outcome. The observer paired with the controller was presented task configurations in the same order but saw the outcome by observing the pen trace produced by the controller.

The experimental results are plotted in ROC form in the four graphs of figure 3.

EXPERIMENT II

The manual control task consisted of driving an automobile between three sets of rubber traffic cones on a parking lot. The spacing and offset of the second two pairs of cones varied randomly from trial to trial. A typical arrangement is shown in figure 4. The driver had to enter the first gate at 20 mph, and decisions or estimates had to be made by this time. Knocking over one or more cones was considered a failure.

There were two subjects in this experiment. On each trial, one served as the controller and the other sat beside him in the front seat as an observer. The controller told his decisions to an experimenter in the rear seat and the observer wrote his down. The observer was prevented from hearing the controller's decisions by earphones presenting white noise.

On half the trials, the controller and observer each chose one of two bets defined by the payoff matrix:

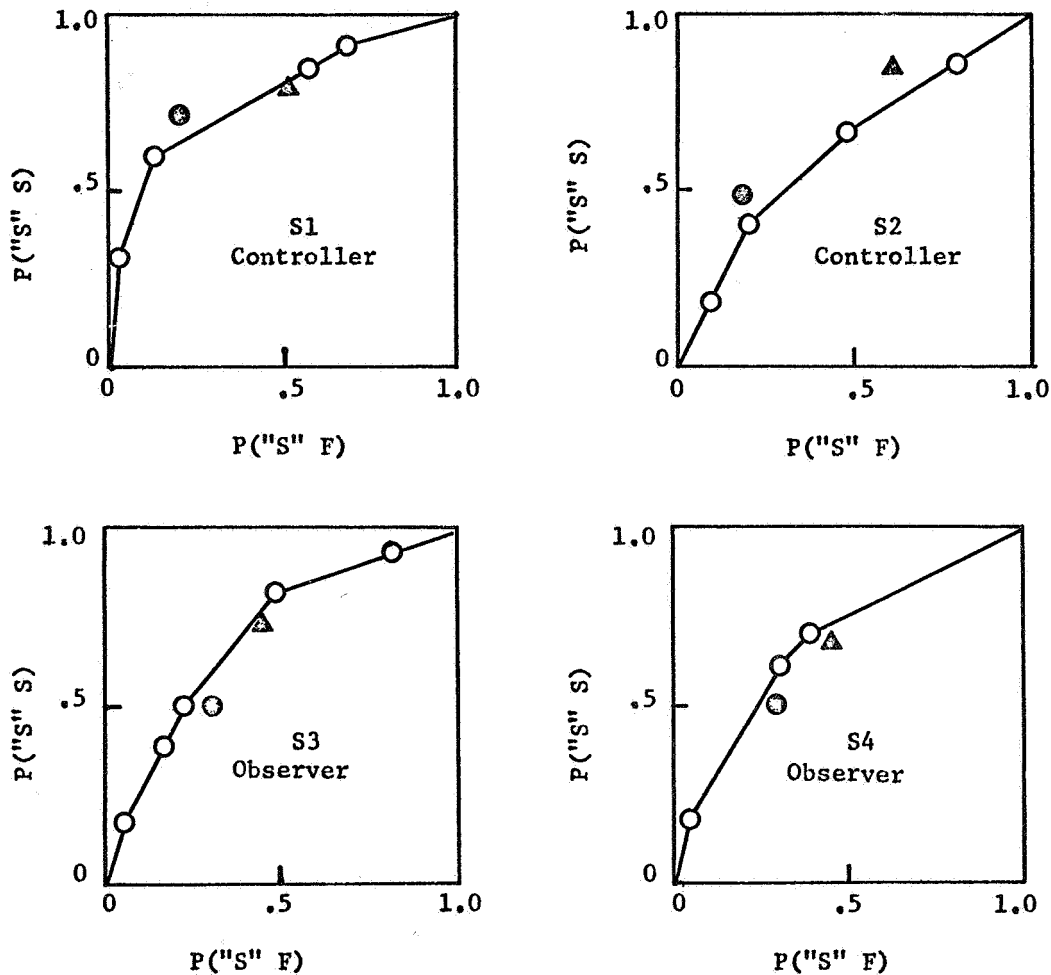
	<u>Success</u>	<u>Failure</u>
Bet A	15¢	-10¢
Bet B	5¢	-5¢

On the remaining trials they each chose a number between 0 and 100 to represent their feelings about the chances of success.

The experimental results are plotted in ROC form in figure 5.

EXPERIMENTAL CONCLUSIONS

In both experiments, the points generated by bet choices fell close to the ROC curve generated by probability estimates indicating that for these tasks, the decisionmaker's predictive ability can be meaningfully described by a single ROC curve.



Bet choices \blacktriangle Bet A \odot Bet B (40 trials/point)
 \circ Generated by probability estimates
 (80 trials/subject)

Figure 3.—Results from experiment I.

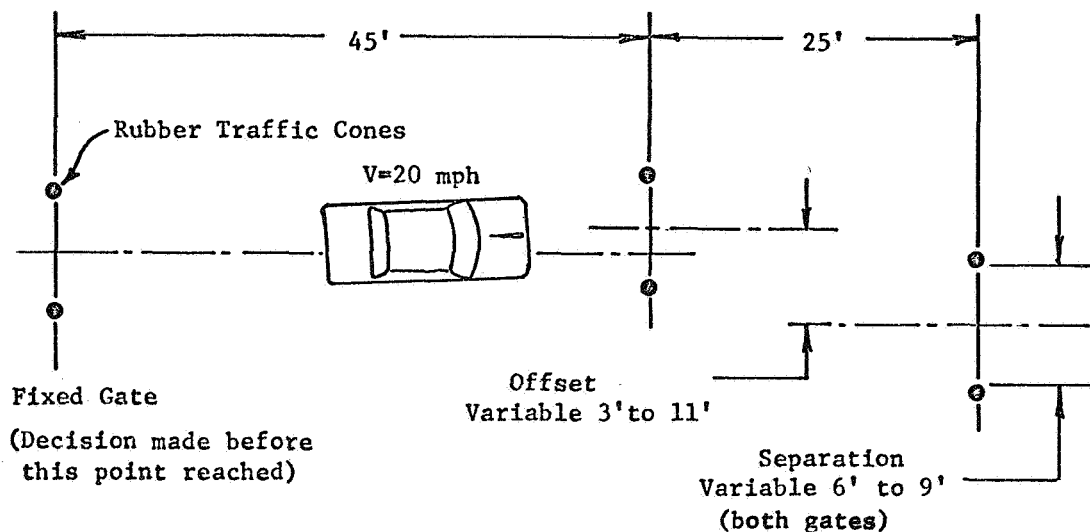


Figure 4. —Task layout for experiment II.

CONTROLLER VERSUS OBSERVER

No significant difference in the shape of the ROC curves of controllers and observers was found. This result indicates that the controller has no further source of information beyond the task configuration and his past performance on which to predict his success. Controllers however tended to locate their operating points above and to the right of observers on their ROC curve.

In terms of the decision model, this means that, though equally able to detect future success, controllers make the choice consistent with success more often. This is much as if they had in addition to monetary payoff an extra, intangible reward for correctly predicting success.

POTENTIAL APPLICATION

VALIDATION OF SIMULATORS.—A simulator is an artificial environment in which events occur analagous to events in the real world. Its effectiveness lies in the strength of the inferences that can be drawn about an operator's behavior in the real world from his behavior in the simulator.

An operator's decisionmaking performance in a simulator will differ from that in an actual situation if either he perceives the causal relations among stimuli differently in the two environments or he has in the two cases entirely different utilities for real and simulated outcomes. The ROC formulation makes it possible to specify precisely the requirements for simulator validity. With respect to the first problem, if the ROC curve for a given task is the same in both environments, then decisionmaking in the simulator for that task is potentially the same as that in the actual situation; for the ROC represents the limiting ability of the decisionmaker to extract relevant information from the data available.

The argument is widely advanced that results from simulators can never be extrapolated with confidence since the operator never risks his life. If it has been established or can be assumed that the same ROC curve holds both in the simulated and real situations, then the effect of a difference in utility for the outcomes under the respective circumstances

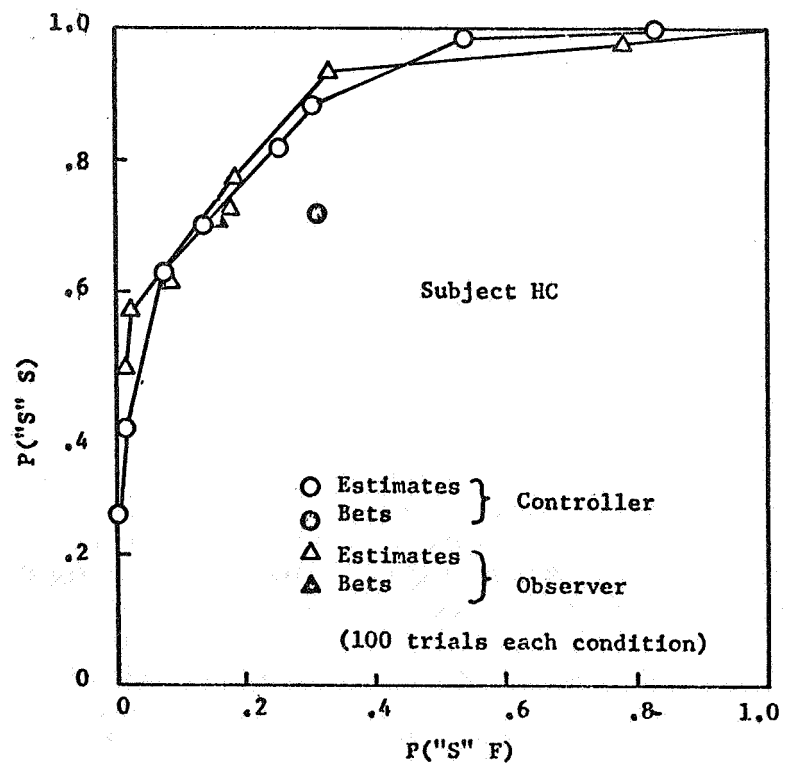
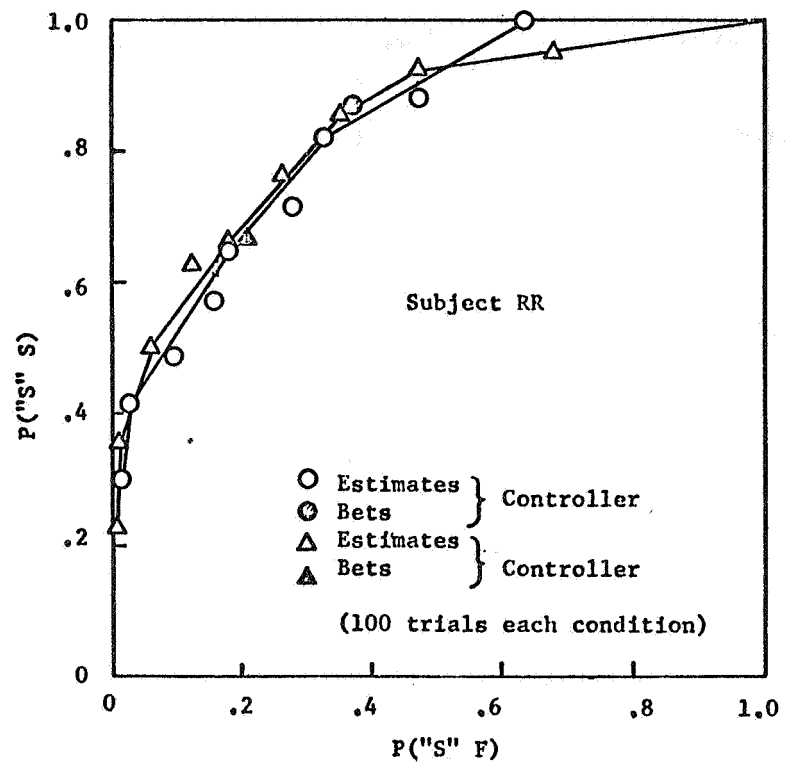


Figure 5.—Results from experiment II.

is solely in the operating point. The greater cost of real failures will result in a more conservative cutoff and the range of operating conditions will be precisely delineated.

EVALUATION OF DISPLAYS. — Man-machine systems sometimes contain displays that augment the man's ability to predict his performance at a control task without influencing his ability to perform the task. These help the man decide whether or not to try a given task. An example is a device that tells a driver how much time remains in the amber phase of a traffic light cycle. This information could help the driver decide whether to stop or try to get through the intersection before the light changes.

The ROC curve is an appropriate measure of predictive ability to be used in the experimental evaluation of such displays, and where necessary enables the separation of the display's enhancement of the operator's control skill from its effect on his predictive ability.

13. The Effect of Psychological Stress on Decision Processes in a Tracking Task

C. B. Gibbs
National Research Council, Ottawa

In a tracking task, moderate alcohol consumption produced impairment that persisted some 7 hours after drinking had ended. Forty hours of sleep deprivation produced impairment but in some subjects, the adverse fatigue effects were offset by moderate drinking. Disturbance of sleep sometimes caused grave impairment up to 1 hour after waking.

INTRODUCTION

The human brain is the most complicated control system on Earth, but little is known of the operational methods that are used within the central nervous system. However, recent research on biological control mechanisms has produced findings and inventions of potential social value in reducing accidents. Many forms of psychological stress impair a person's ability to control a machine and are undoubtedly major causes of accidents.

Stress is defined as a force that can produce strain or distortion in a system. Psychological stress can be applied to human beings by the use of alcohol, drugs, or by fatigue, frustration, or many other influences. It is extremely difficult to detect impairment in actual performance because increased expenditure of energy or effort may mask the effects of stress on the aspects of behavior that are readily observed and measured.

The precision of tracking movements made in a limited time is frequently measured in laboratory studies of stress. Precision is obviously important in controlling machines, but it is clear that incorrect decisions, rather than lack of precision in executing movements, are major causes of accidents. Reference 1 describes a tracking task wherein subjects had a choice between rapid decisions, which usually produced incorrect responses, and more deliberate judgments, which normally led to appropriate action. Reference 2 shows that the rate of making accurate decisions is often impaired by moderate doses of alcohol, and there can be no serious doubt that impaired judgment is a major cause of the high correlation between drinking and accidents that is described in reference 3.

THE STRESSALYZER

There is a serious shortage of rapid, sensitive tests of the effects of stress on those aspects of skill that are known to be important in controlling vehicles. In an effort to provide more sensitive measures, an instrument called the stressalyzer was developed to provide a rapid and simple test of a person's ability to control a machine. A handwheel and

pointer are used to track a target that appears equally frequently, at 2-second intervals, at one of five different positions. The handwheel and the pointer move in opposite directions, and this arrangement causes the operator to make an appreciable number of tracking errors, i.e., tracking movements that start in the wrong direction.

A target starting from either of the outside positions 1 or 5 can only move inward, and these steps are termed "unequivocal." There is a 3 to 1 probability that a target starting from position 2 will next move to the right rather than outward to position 1; a target at position 4 is more likely to move left than right. Steps are termed "probable" when the movement actually demanded conforms to the higher probability; steps outward from positions 2 or 4 are called "improbable." A target starting from position 3 is equally likely to move in either direction, and these steps are called "equiprobable."

Response latency or reaction time is the period elapsing between the appearance of a target and the beginning of the operator's response. The stressalyzer provides simple, readout records of response latency, the number of response latencies exceeding 0.4 second, the number of errors, and the number of targets acquired with high precision in one run.

Figure 1 shows the average results of a group of 12 subjects. The height of the solid black bars indicates the percentage of errors on improbable and probable steps, and it is clear that error percentage is far higher on the former than on the latter steps. The

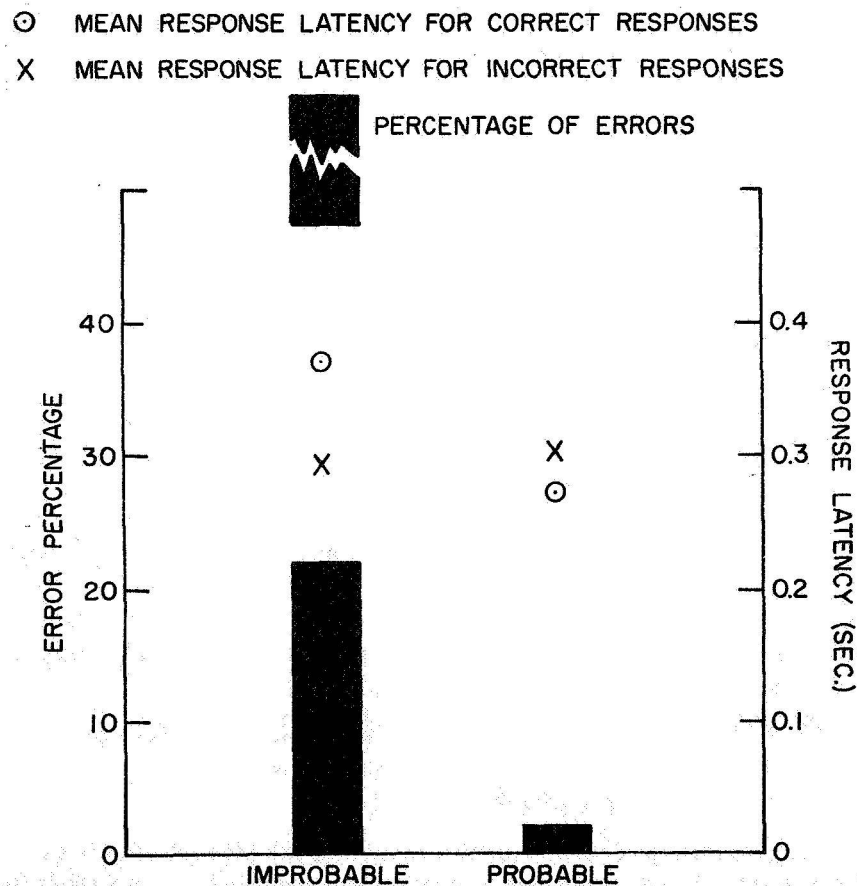


Figure 1.—Mean response latency for correct and incorrect responses and percentage of errors on steps of different probability.

height of the circles represents response latency for correct responses, whereas the latency of movements that start in the wrong direction is indicated by the height of the crosses. The graph shows that it takes much longer to reach a correct decision on improbable than on probable steps. On improbable steps, short latency usually produces an error, whereas more deliberate decisions normally lead to a correct response.

THE EFFECTS OF SLEEP DEPRIVATION ON SKILL

A tired subject is often stimulated by the challenge of a test situation and may perform well for a period. Performance on some tasks is impaired little, if at all, by prolonged sleep loss. Sleep deprivation reduces efficiency in a typical vigilance task of detecting a faint, infrequent signal, but even when subjects have normal sleep, their performance often deteriorates as their time on watch increases. In many tasks, it is difficult or impossible to decide whether sleep loss reduces motivation through the effects of boredom or whether it produces actual impairment of skill. The theoretical issue has important practical implications because reduced motivation can be offset by a variety of means, whereas the condition of impaired skill can only be remedied by rest and sleep.

Previous studies of sleep deprivation, using conventional methods of testing, have produced a variety of opinions on the causes and effects of impairment by sleep loss. Reference 4 indicates that 30 hours sleep deprivation had little effect on performance when subjects were given knowledge of the results of their performance, but there were adverse results when this information was withheld. Wilkinson (ref. 5) suggested that motivational factors were important determinants of the degree of impairment, and concluded that "Sleep deprivation reduces the level of arousal or activation of the body. If the incentive is there, however, this can be overcome, so that arousal and therefore performance are maintained at normal levels."

There is an urgent requirement for a rapid test that can detect impairment of skill by fatigue, as distinct from the boredom that may be produced by a lengthy test. The stressalyzer sets a task of brief duration and provides measures of response latency, errors, and the number of response latencies exceeding 0.4 second. These measures relate to the speed and precision of movements and the rate of making correct decisions on improbable events; the long latencies may well reflect or precede potentially dangerous lapses of attention. These are important factors in controlling vehicles and other machines, and the stressalyzer was therefore used in studies of sleep deprivation.

The very different effects of fatigue on various tasks may be due to differences in their inherent interest, as suggested in reference 5, or to distinctions drawn in references 6 and 7 between the times available to complete different tasks. The stressalyzer presented a paced task wherein targets appeared for a limited time. An unpaced task of mirror tracing was also used in the studies of sleep deprivation. The subjects used a metal stylus to trace over the outline of a star while viewing reflections of the star and stylus in a mirror. The total time taken for tracing and the period wherein tracing was precise were both recorded.

SLEEP DEPRIVATION FOR 48 HOURS

Twelve youths, aged between 16 and 19 years, volunteered for the tests and were given preliminary, spaced practice totaling nearly 4 hours on each task. They were then

deprived of sleep for 48 hours, while kept under close supervision. The experiment started at 0900 hours and each subject was tested on both tasks in each subsequent 4-hour period. The group had three runs on the stressalyzer in each test session, and a similar time was spent on mirror tracing.

The average scores of the group on three successive runs on the stressalyzer are shown in figure 2, wherein captions define the meaning of the plotted symbols. The figure shows separately the mean response latency (RL), the number of targets missed (TM), and directional errors (E). The combined effects of these different effects are also shown by a statistic based on the product of response latency and the sum of targets missed and errors. The actual statistic used, $4(RL)(TM + E)$, was derived to provide sensitive discrimination among the abilities of different subjects, among the effects of different experimental conditions, and to provide a suitable scale for the graphs.

The figures at the foot of figure 2 show the hours elapsed from the start of the test. There was an appreciable rise in the statistic after sleep deprivation of 20 hours. There was also a rise in the percentage of long response latencies exceeding 0.4 second, denoted by closed circles, some 10 hours after starting the test, with another pronounced rise when the group had been deprived of sleep for 28 hours. The detailed results show large differences among individuals in their tolerance of sleep deprivation.

Figure 3 shows the results of subject 3; the figures at the top of the graph show the time of day and, immediately below this, the relative ranking of the individual in a group of 12 subjects, in respect to the various measures that were obtained. In overall ability, as shown by the statistic, subject 3 began with the worst score in the group but ended the test ranking third because he maintained a more consistent performance than many other subjects although there was no improvement in his absolute scores.

Figure 4 shows the results of subject 2 and the considerable deterioration in his overall ability, as denoted by the statistic, only 10 hours after testing began, although his initial ranking was higher than that of subject 3. There was also a large increase in his percentage of long latencies some 16 hours after testing began.

Figure 5 shows the results of subject 12 who ranked seventh in overall ability, as denoted by the statistic, when the tests began, but ranked only tenth after 48 hours without sleep. Comparison of figures 3 and 5 shows a hare and tortoise effect, because the initial ranking of subject 3 with respect to the statistic was lower than the ranking of subject 12, but the order was reversed by the end of the test.

Figure 6 shows large fluctuations in the percentage of long latencies produced by subject 8. This percentage was extremely high between 0900 and 1300, some 28 hours after beginning the test, but at this stage his movements were still rapid and precise, few targets were missed, and few errors were made. In tests conducted between 0100 and 0500 hours, with sleep deprivation of about 44 hours, there was again a high percentage of long response latencies accompanied by considerable impairment of tracking ability.

Sleep deprivation for 48 hours had much less effect on the unpaced task of mirror tracing than on the paced task set by the stressalyzer. In tracing, the group's percentage time on target was 81 percent in the first test session, with a slight drop in precision to 77.6 percent in the last session. In contrast, the average time per run was reduced from 37 seconds to 33 seconds in the first and last sessions, respectively.

THE EFFECT OF DISTURBED SLEEP ON SKILL

Reference 5 notes the paucity of studies of partial sleep deprivation, which is probably of greater practical importance than complete deprivation. The previous study was therefore extended. The same 12 youths were again tested over a 48-hour period wherein sleep

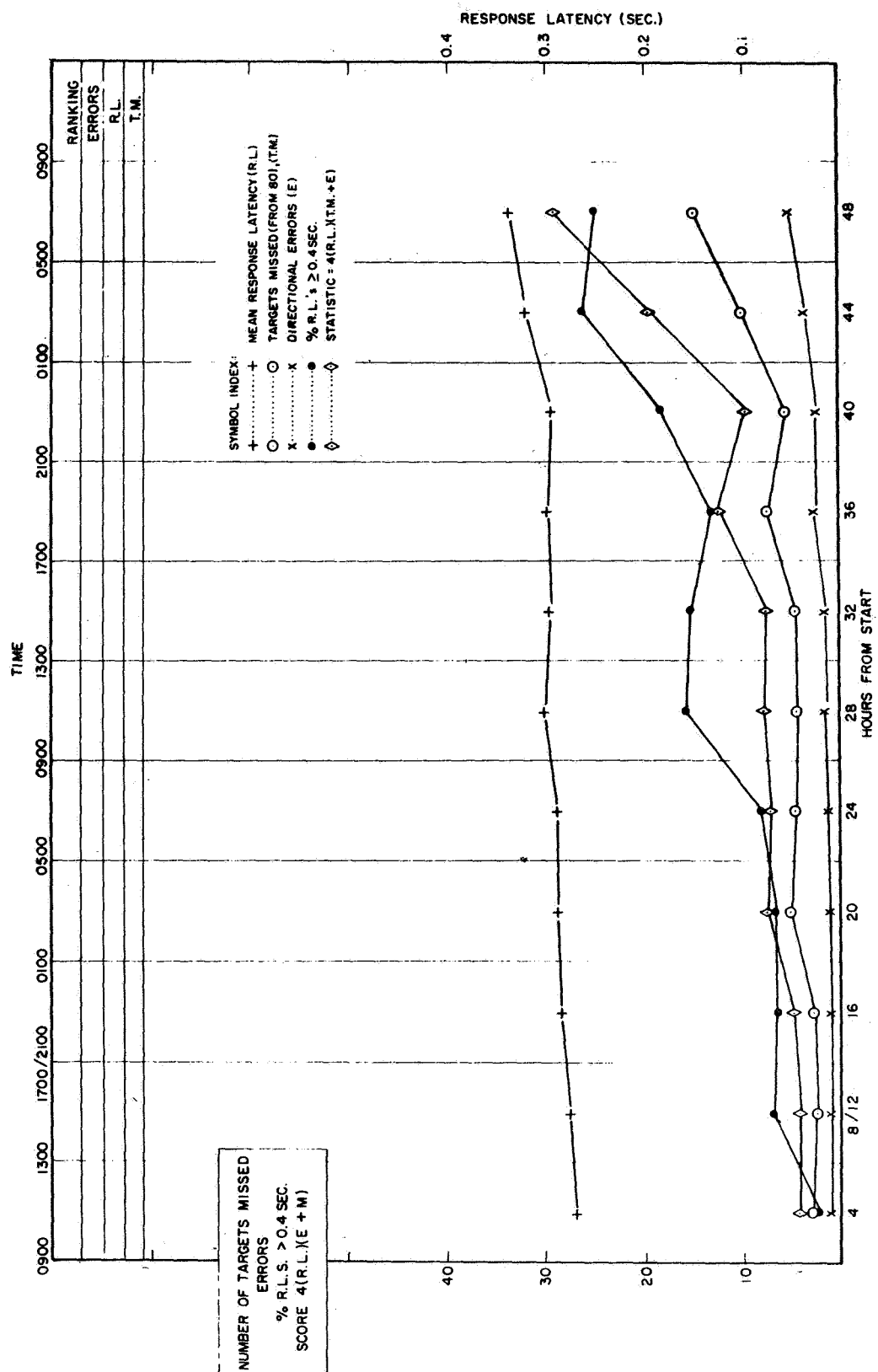


Figure 2.—Group results, July 15, 1966.

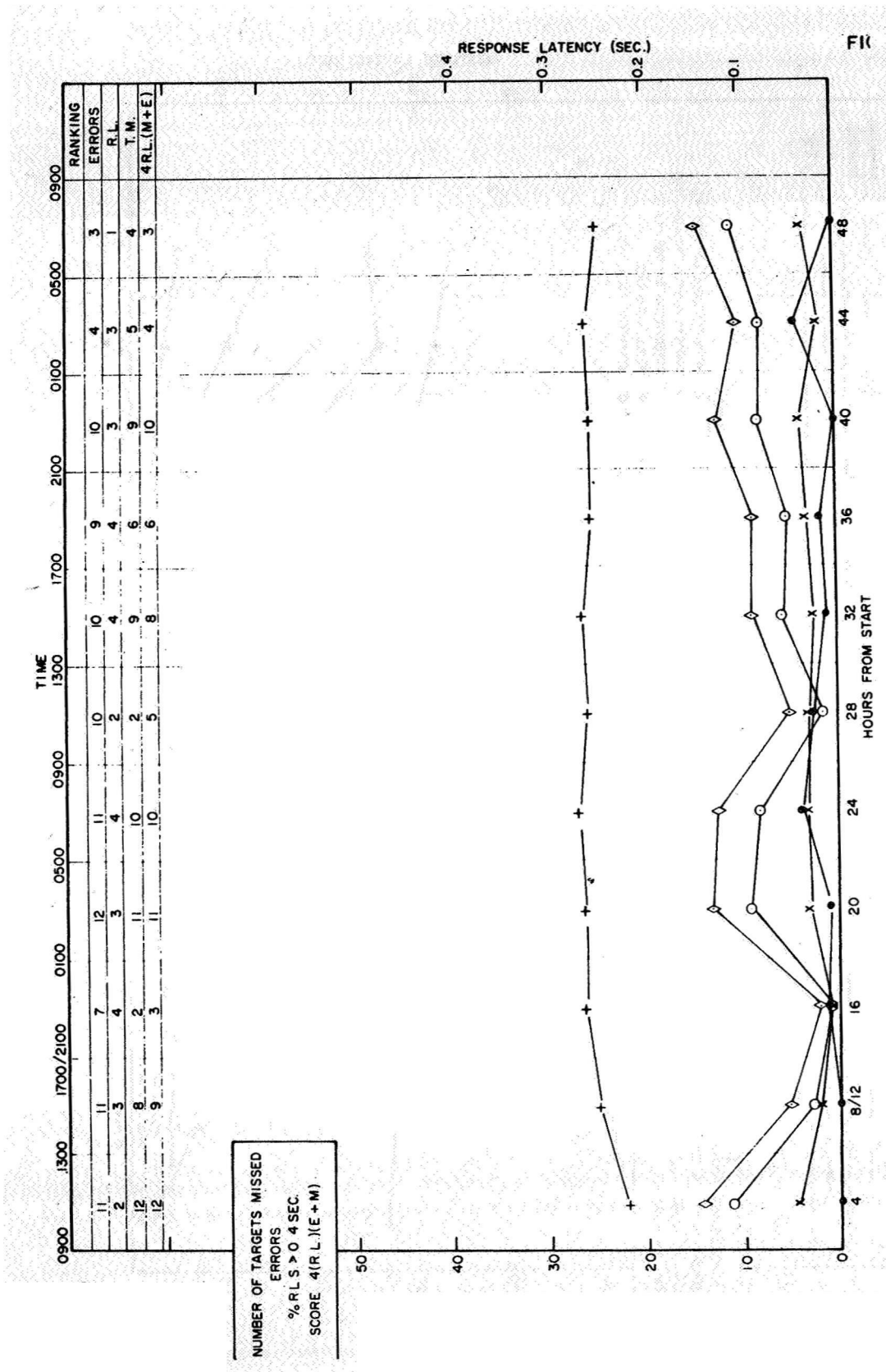


Figure 3. —Subject 3.

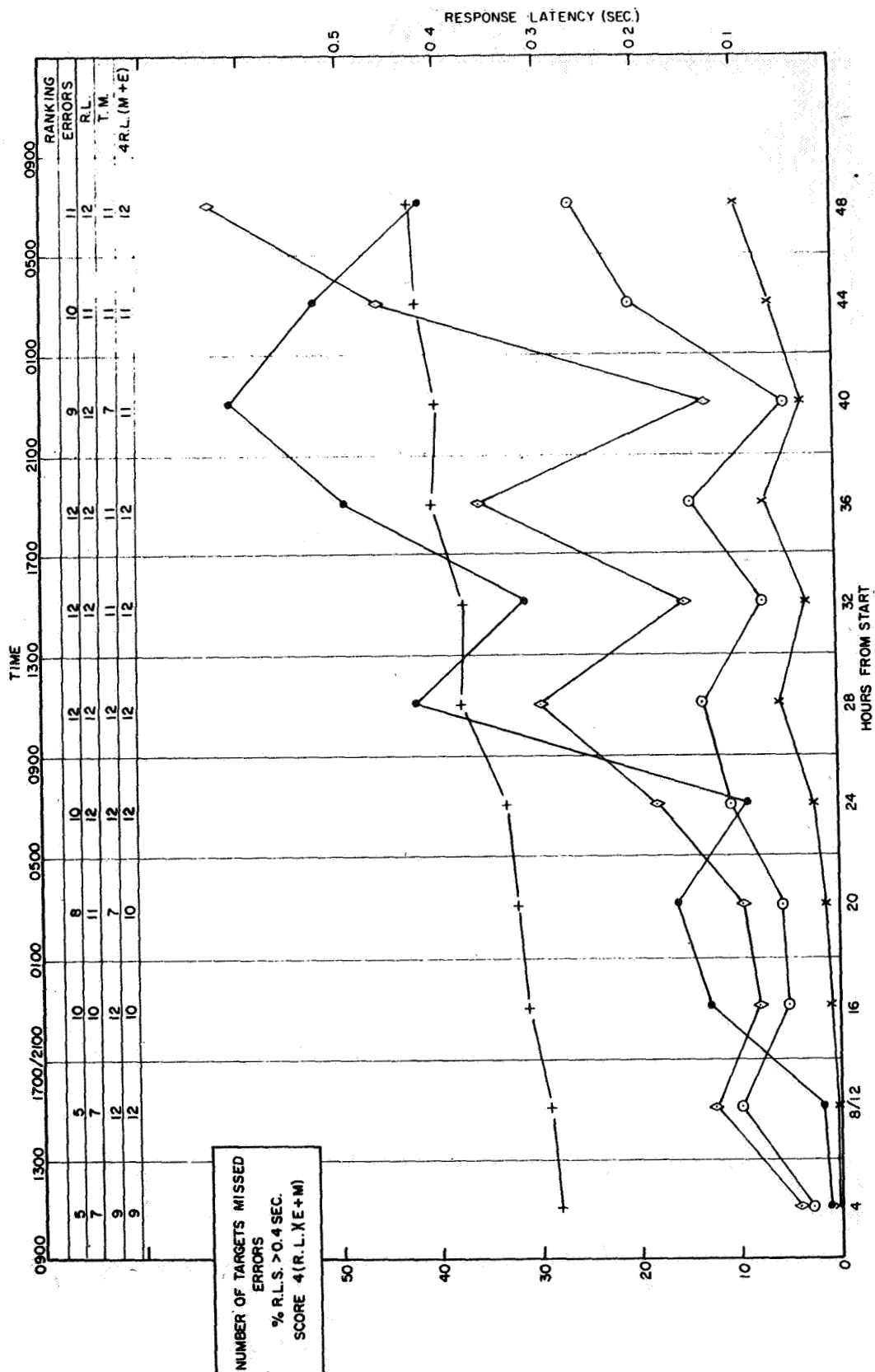


Figure 4.—Subject 2.

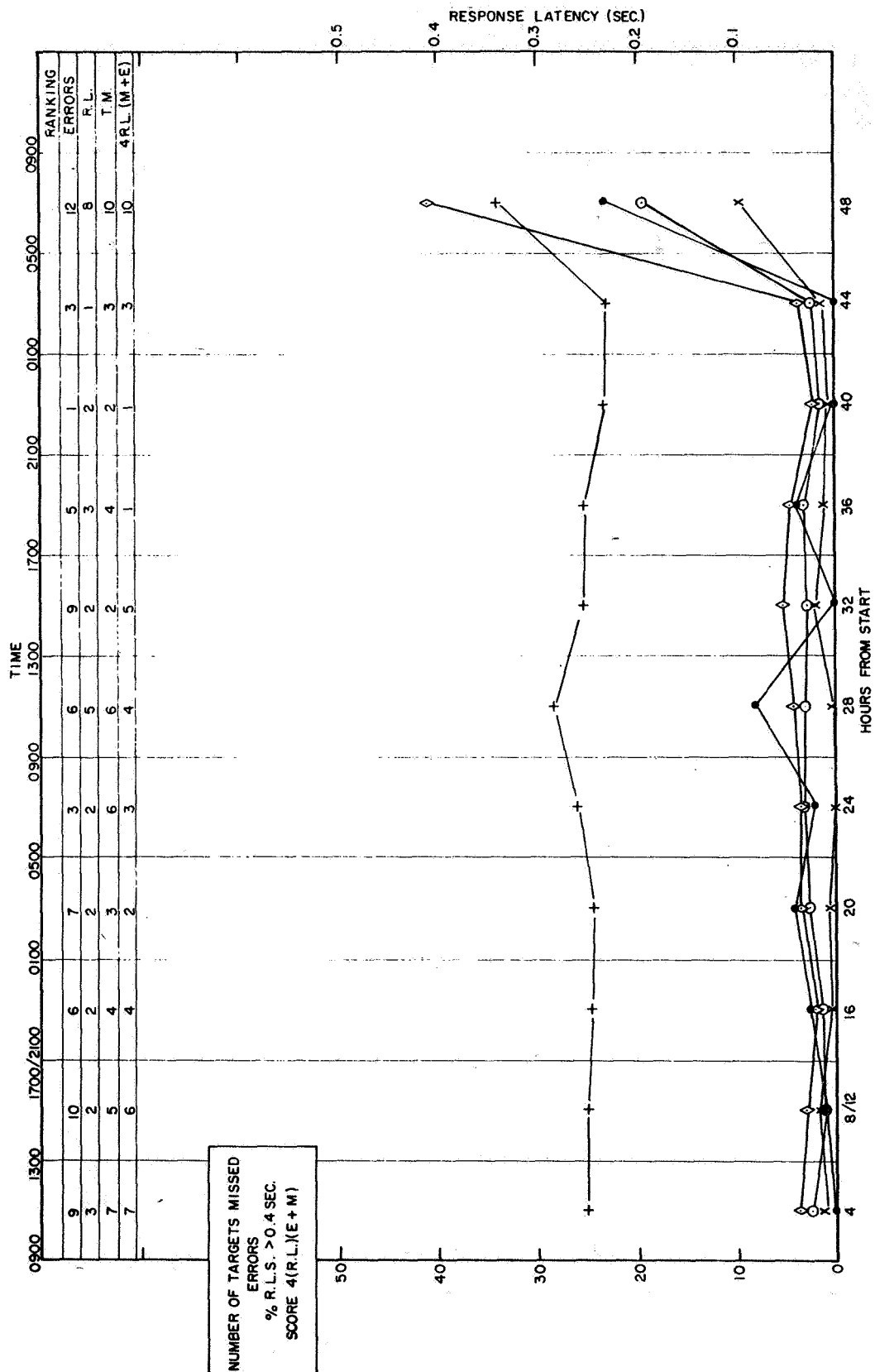


Figure 5.—Subject 12.

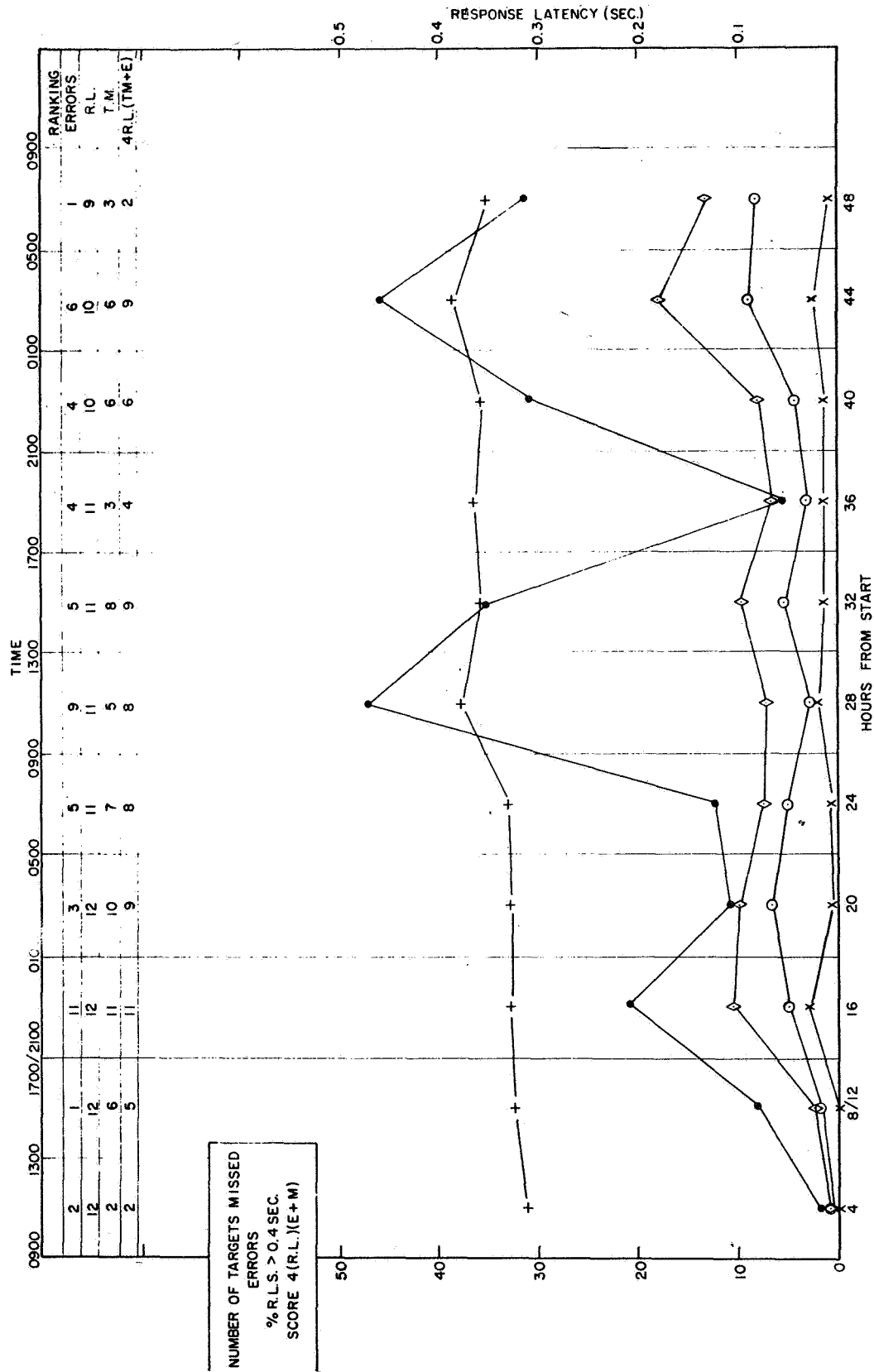


Figure 6.—Subject 8.

was permitted, although sleeping subjects were aroused periodically at night and tested some 15 minutes after waking, at the times and intervals used in the previous sleep-deprivation experiment.

Figure 7 presents group data for the mean of three runs made during each test session on the stressalyzer, and shows considerable deterioration in group performance in the early morning, between 0100 and 0900 hours. Figure 8 shows the detailed scores of subject 2 and allows comparisons with his scores shown in figure 6, which were obtained in the study of complete sleep deprivation. The impairment of subject 2, soon after waking in the early morning, exceeded the deterioration produced in any one subject by 48 hours complete sleep deprivation in the previous study, or by the highest alcohol dosage tested in studies that are described later. Another subject was found to be impaired, to an even greater degree than subject 2, when tested soon after waking.

THE COMBINED EFFECTS OF SLEEP LOSS AND ALCOHOL

Eight young men in another group were kept awake for 36 hours and had three runs on the stressalyzer at five hourly intervals. The subjects drank spirits before taking the last test and the average breathalyzer reading of the group, in the last test session, was about 0.08 percent.

A surprising finding is illustrated by figure 9. In spite of intoxication and sleep deprivation, the average group scores were better on the last run than on the first. Five of the eight subjects had worse scores on their first run than on their last trial. Figure 10 shows the scores of subject H, who was one of two subjects exhibiting gross impairment in early trials, although all subjects had been awake for at least 1 hour before their first test began. The scores of this subject and of the group as a whole improved some 5 hours after testing began, and from then on increasing sleep deprivation, together with alcohol consumption before the last session, produced little or no impairment in respect to response latency, targets missed, or errors. Figures 9 and 10 show a high percentage of long latencies, however, during the last session.

A COLLABORATIVE STUDY

A further study was conducted in collaboration with the Human Resources Research Office (HumRRO) of the George Washington University. The HumRRO scientists conducted tests on a simulated driving task, and a task of monitoring four screens in a typical vigilance test in which the subjects were required to report the appearance of a faint light on a screen. The National Research Council (NRC) team conducted tests on the stressalyzer, and the two groups communicated data on a reciprocal basis.

In the HumRRO study, five test booths were provided and two men occupied each booth. One man, termed the driver, used a steering wheel to position a tracking index, in a simulated task of steering along a winding road. The scores termed "time off target" denote the period wherein the tracking index was outside the limits of the simulated road during a test session of 90 minutes. The other man, termed the monitor, watched four large screens and was required to report the appearance of a faint orange light that appeared briefly, at random intervals, on 15 occasions during a test period of 90 minutes. Both tasks, of monitoring and simulated driving, were carried out continuously during each test session. A test session was normally followed by a rest pause of 15 minutes, but there was a longer break of 1 hour for a meal, beginning at 0700, 1300, 1900, and 0100 hours. Each man had three successive runs on the stressalyzer at intervals of about 4 hours. The times of testing and the average results of 10 men who were tested are shown in figure 11.

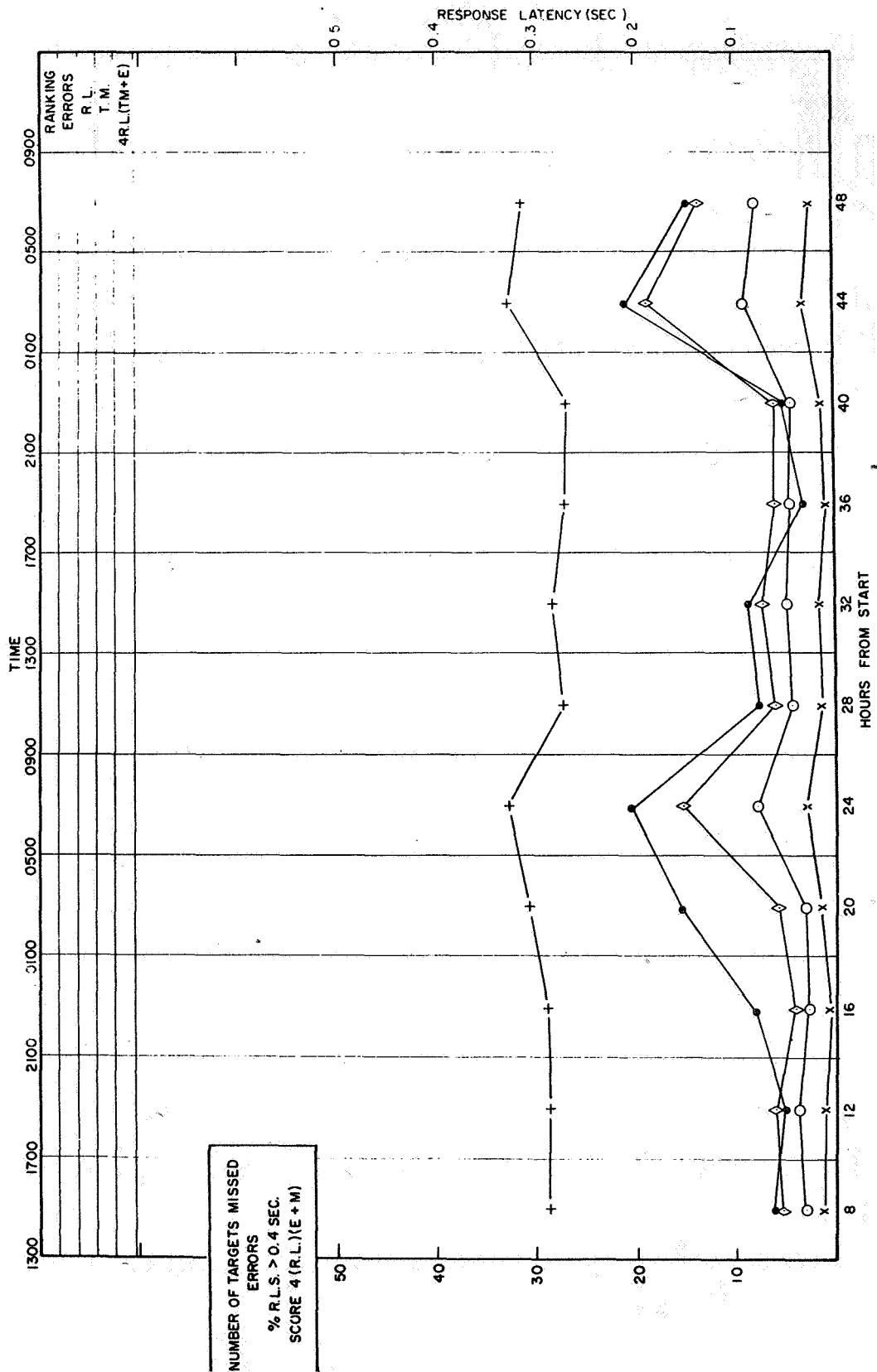


Figure 7.—Group results, July 22, 1966.

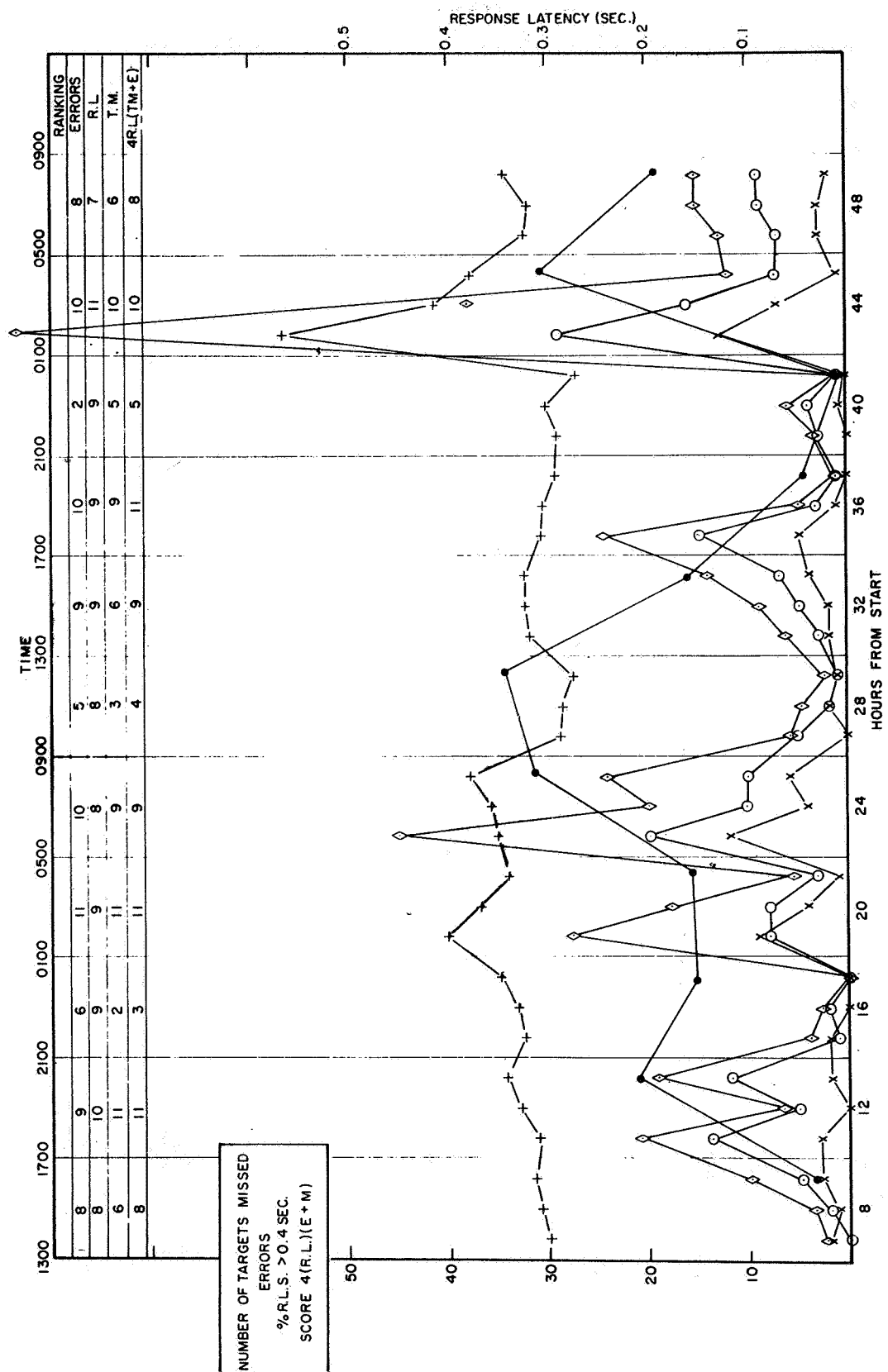


Figure 8.—Subject 2.

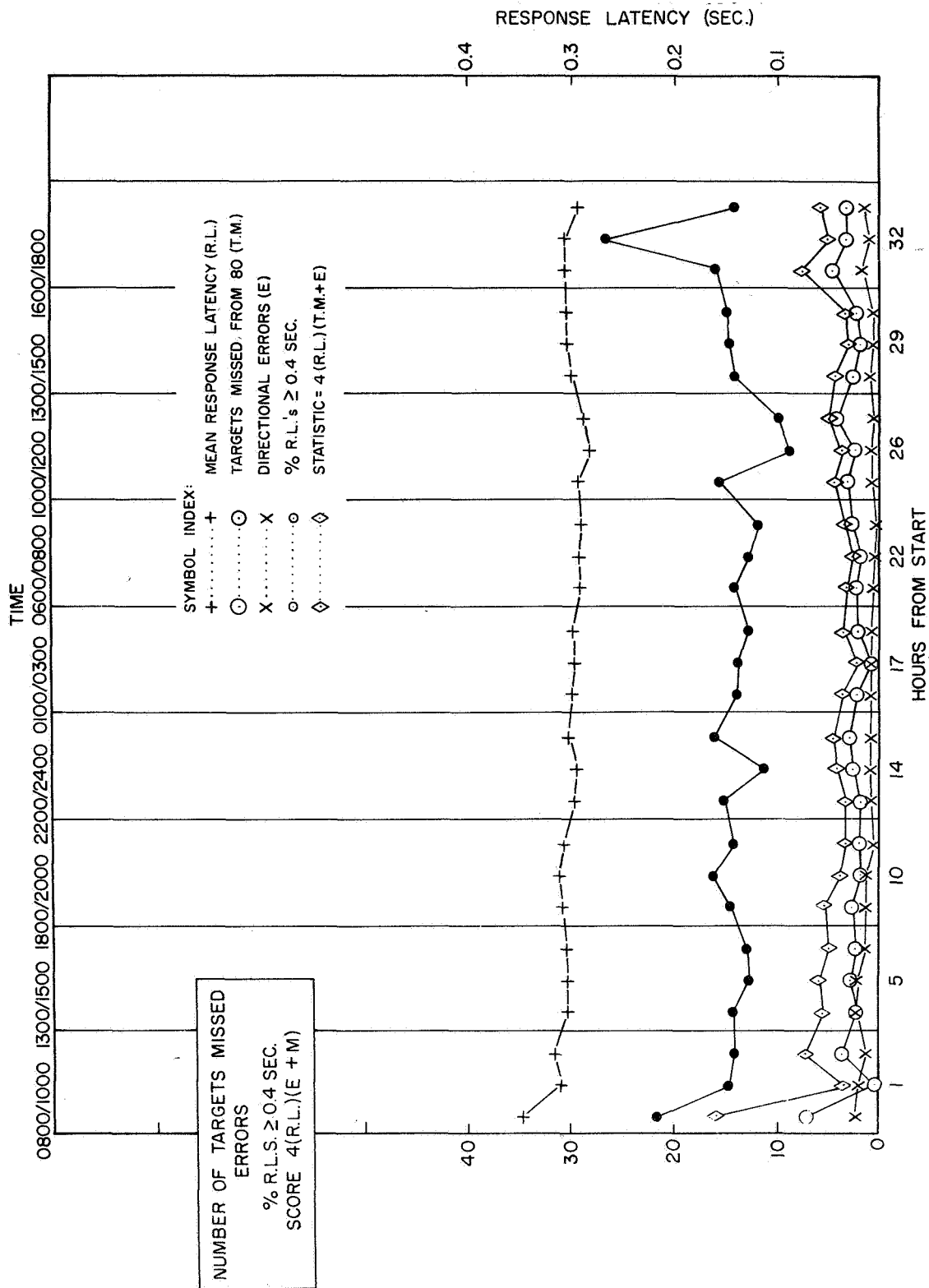


Figure 9. —Group results, 36-hour test.

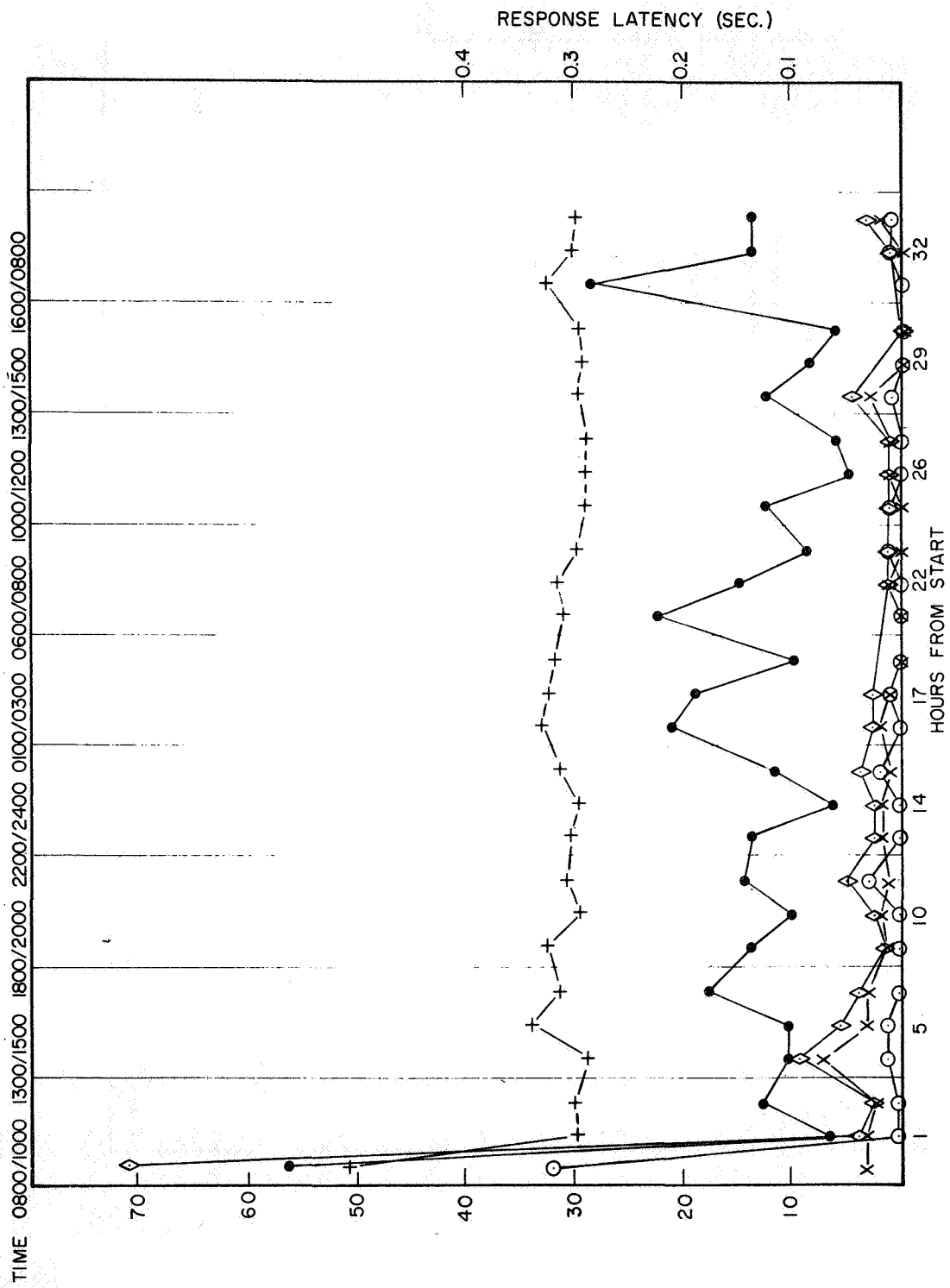


Figure 10.—Subject H.

Figure 11 shows the group's average results on the stressalyzer, using symbols already defined. Squares are used in plotting the data kindly provided by HumRRO scientists, and again a rise in the curve indicates worsened performance. The solid squares represent the monitors' scores in a 90-minute session. The number of signals missed is expressed as a proportion of the total number of signals presented. The open squares indicate the drivers' performance; the time off target is expressed as a proportion of the total 90 minutes of the test.

Figure 11 shows a large increase in the proportion of signals missed, about 21 hours after starting the test, but the drivers did not show a comparable increase in time off target until near the 33-hour stage. Over much of the test, the signals missed on the vigilance task and the latencies of 0.4 second or more on the stressalyzer, tended to increase or to decrease together. The results of subject F, shown in figure 12, exemplify this correlation.

THE EFFECTS OF ALCOHOL ON SKILL

Research on the effects of alcohol stress has been hampered by the lack of an adequate analysis of skill and the absence of rapid and relevant test procedures that could be standardized. Chemical and physiological tests are available that show the effects of alcohol on the breath or blood of subjects, but such tests do not provide insight into the effects of alcohol on actual behavior.

Figure 13 shows the effects of alcohol on three individuals, selected from 150 subjects tested to date, to illustrate the wide range of differences among individuals in their innate ability, tolerance of stress, and the strategies they adopt under alcoholic stress. The number of errors is shown by triangles scaled on the left ordinate, and response latency is indicated by circles, scaled in portions of 1 second, on the right-hand ordinate. The errors and response latencies are plotted separately for each subject when his breathalyzer reading was zero, and after drinking had raised the breathalyzer level to about 0.05 percent and then to 0.1 percent or more, as shown at the foot of the figure. The first graph shows the average results of three runs by a subject who progressively reduced his response latency with increased intoxication, with a consequent large increase in errors. The third graph is from a subject with a similar preference for rapid reactions rather than slow, accurate decisions. The second graph is from a subject who adopted the opposite strategy by increasing his reaction time as intoxication increased, thereby minimizing his errors. This subject made fewer errors with a breathalyzer reading of 0.11 percent than the other subjects made when they were sober.

Figure 14 shows the average results of a group of 12 subjects who were first tested when sober, and later after drinking had raised the group's average breathalyzer reading to 0.05 percent and then to 0.1 percent. All subjects had three runs on the stressalyzer at all three breathalyzer readings. Drinking ceased on reaching the 0.1 percent level, but the subjects were tested later when the average breathalyzer reading of the group had fallen to 0.05 percent and then to zero. The two graphs of figure 14 show results on two consecutive days, and it is clear that increasing dosage did not increase errors on highly probable steps, whereas there was a large increase of errors on improbable steps. In figure 14, the circles show average response latency on improbable steps, and the graphs indicate considerable increase in latency with increased dosage. A striking finding is that impairment persisted some 7 hours after consuming a dosage that produced a peak reading of about 0.1 percent, although breathalyzer readings had then fallen to zero. There was little difference between this residual impairment and the deterioration shown with a breathalyzer reading of 0.5 percent.

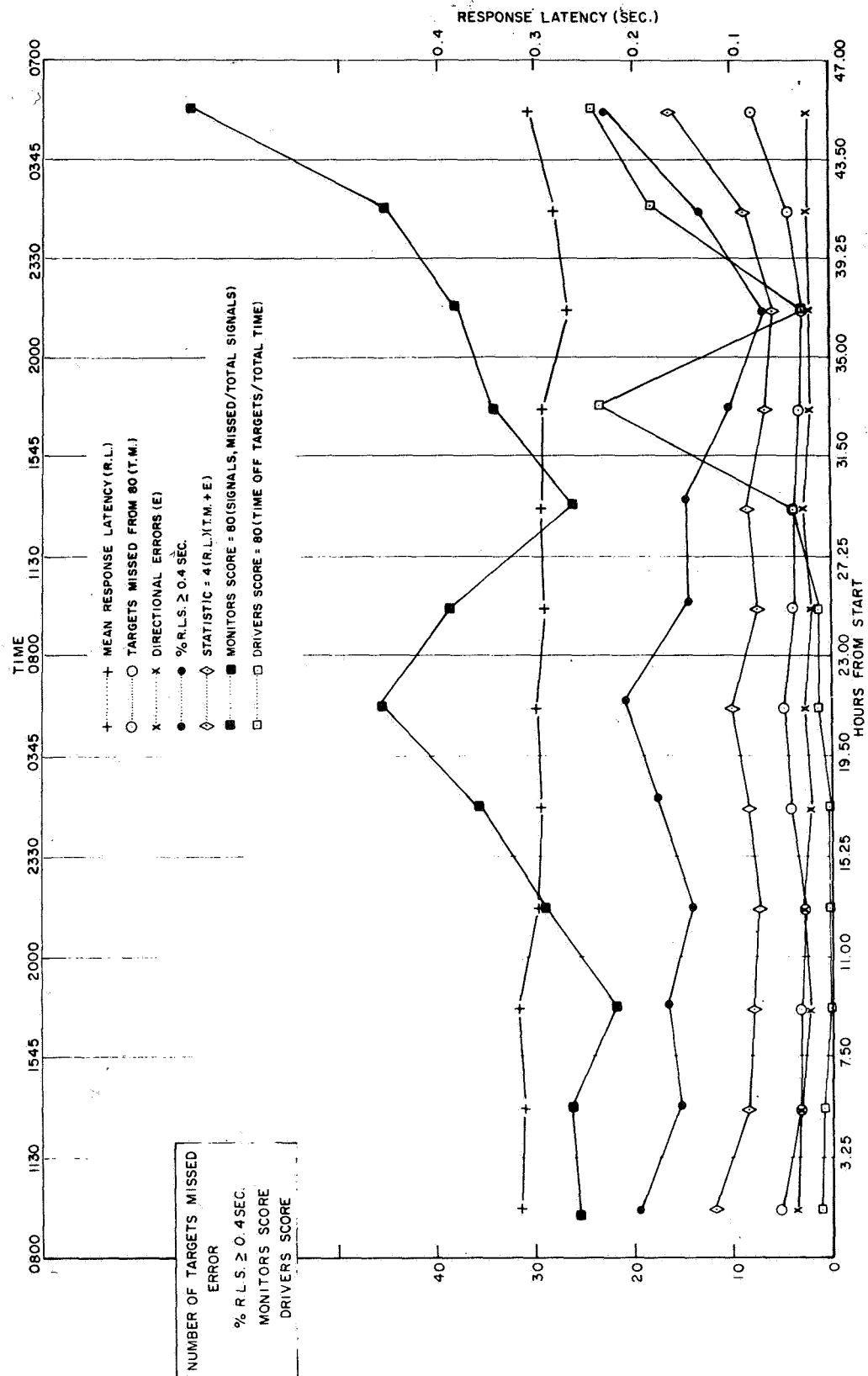


Figure 11.—U.S.A., group I, results.

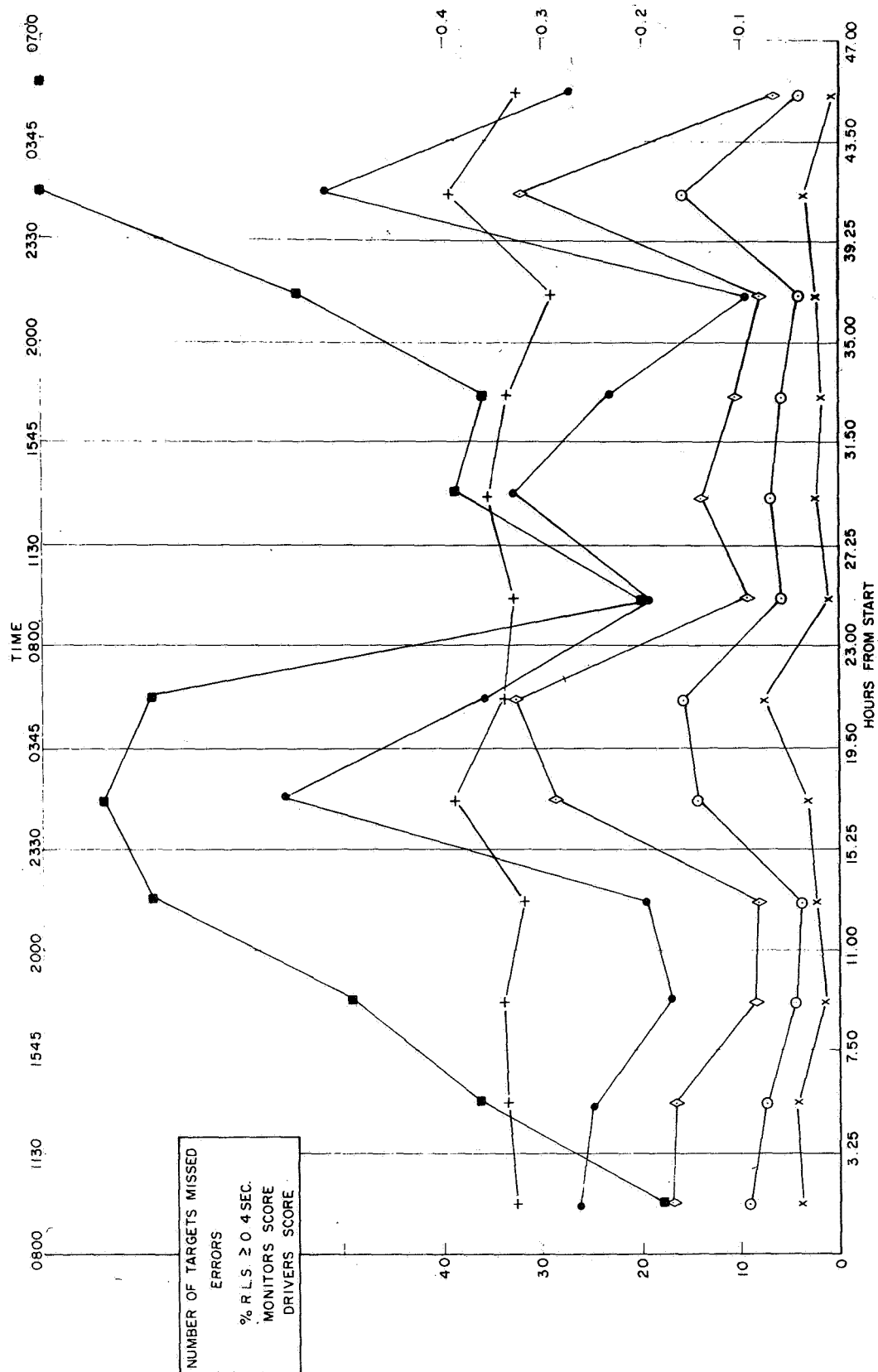


Figure 12. --Subject F, U.S.A., group I, monitor.

○ RESPONSE LATENCY
 △ ERRORS

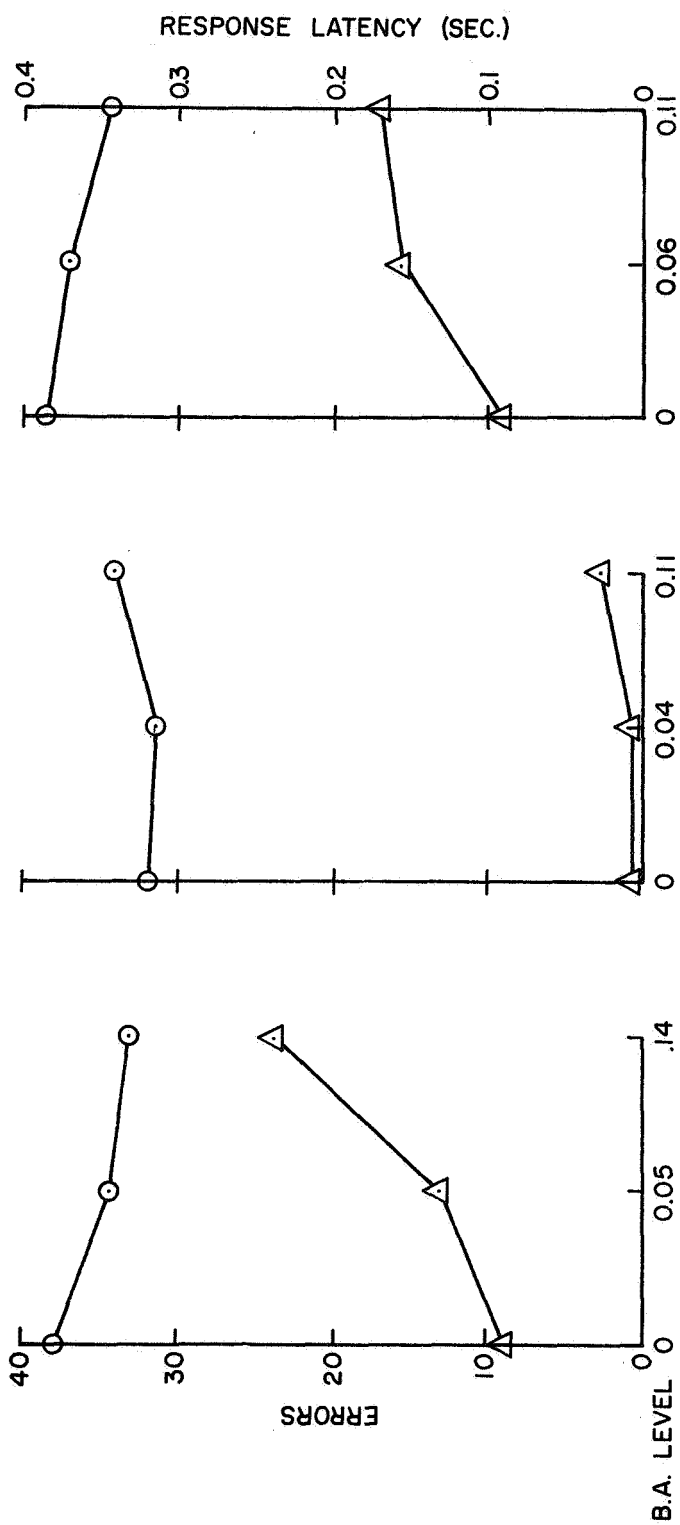


Figure 13. —Individual reaction to alcohol.

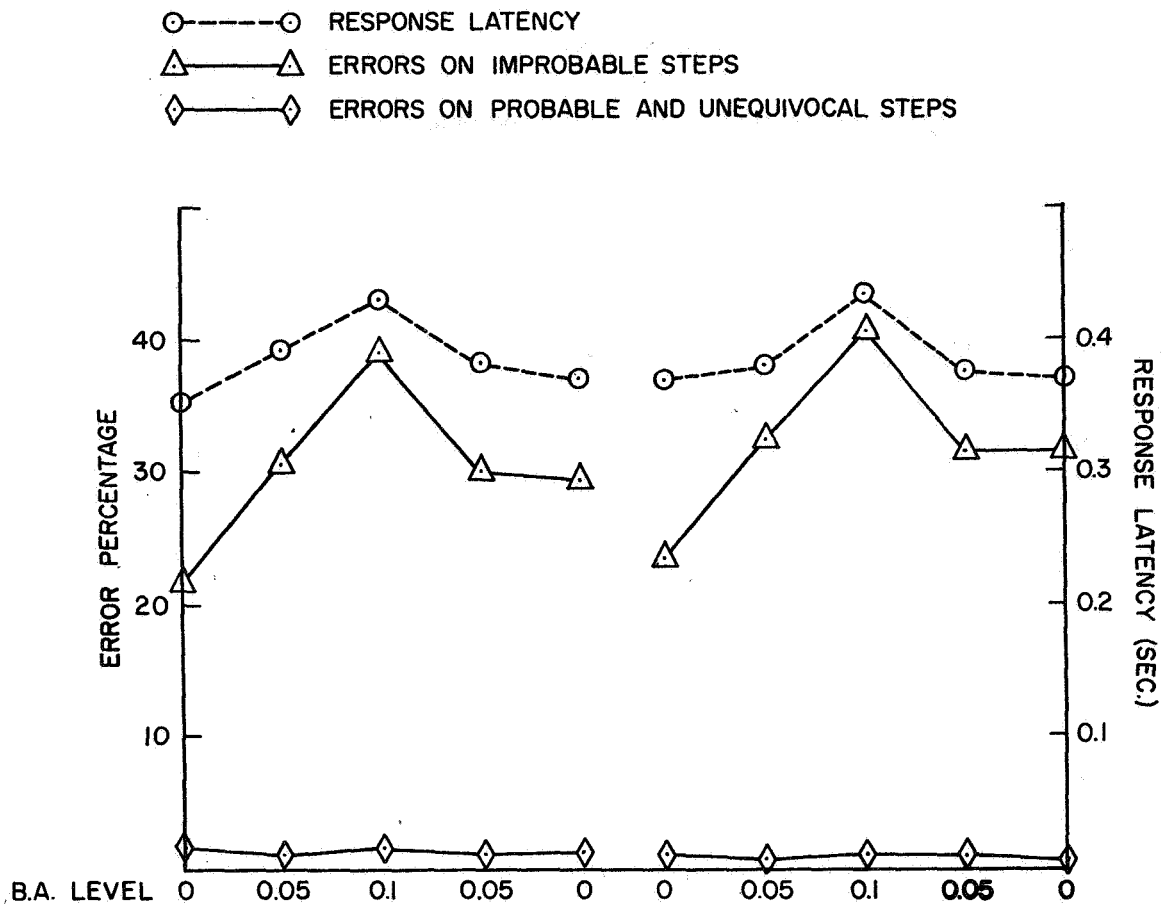


Figure 14. —Effect of alcohol on response latency and errors on probable and improbable steps.

DISCUSSION

The stressalyzer sets a task that was first developed to analyze perceptual-motor skill, rather than to study stress. In an early study, three of four subjects showed gross impairment compared with their previous performance. These impaired subjects admitted to heavy overnight drinking, although this had ended 10 hours before testing began. These subjects were not given knowledge of results and their performance deteriorated steadily over the test session of 30 minutes.

Another group of 12 subjects slept in the laboratory to permit strict control of drinking. These subjects were tested on the stressalyzer but were not told their scores. Figure 14 shows considerable residual impairment many hours after drinking had ended and breathalyzer readings had returned to zero, although figure 2 shows no comparable increase in the errors of sober subjects in the first 20 hours of testing. In the alcohol studies, at the zero breathalyzer level, 11 of the 12 subjects performed better before drinking began than after drinking

had ended; the other subject had equal scores on predrinking and postdrinking runs. This study leaves no doubt of the highly adverse aftereffects of alcohol, in the experimental conditions.

In all other experiments, subjects were given knowledge of their results and separate tests showed that such information sometimes improved performance and tolerance of stress, compared with test conditions when no knowledge was given. The largest change was that one subject acquired 90 of 100 targets presented when given knowledge of results, but only 58 on his second run after this information was withheld.

Figures 3, 4, and 5 illustrate large individual differences in innate ability and tolerance of sleep deprivation, similar to the variation shown in the alcohol studies. Of the 12 subjects, two showed very high resistance to fatigue and little impairment at the end of 48 hours of sleep deprivation. Figure 4 shows that subject 2 was definitely impaired 8 to 12 hours after testing began, and one other subject also showed very low tolerance of stress. The remaining eight subjects showed intermediate tolerance, and figure 5 shows representative results of good performance until the 44-hour stage but gross impairment in the final test session.

Figure 7 shows the results of the same group in conditions where subjects slept at night but were awakened in turn at about 4-hour intervals through the night and tested 15 to 20 minutes after waking. A comparison at the 24-hour stage, based on the statistic 4 (RL) (TM + E) of figures 2 and 7 shows that the group performed very much better in the condition of complete sleep deprivation than in the test where they slept but were tested soon after waking. A similar comparison at the 44-hour stage shows that impairment from disturbed sleep had a cumulative effect that was certainly not less than the impairment following 44 hours of complete loss of sleep. Figures 4 and 8 show the scores of subject 2, and similar comparisons, using the statistic, support the above conclusion. The scores of subject 2 at the 44-hour stage (fig. 8) were far worse than the results of any subject in the studies of alcohol effects and complete sleep deprivation.

A further group of eight subjects had no sleep for 36 hours, and consumed alcohol before the last test session began. In the interests of realism, the overnight drinking of these subjects was not controlled, but they were asked to state the amount of alcohol consumed on the night before testing. Cross checks were made within the group. The results of the group are shown in figure 9, and there was little apparent impairment during their last run, compared with their first. In everyday observation, two sources of stress may tend to cancel rather than compound each other, as in the familiar example of loud noise tending to offset the effects of sleep deprivation. In the present study, the consumption of a limited amount of alcohol did not have the expected effect of lowering group performance, but this result may be due to the introduction of a novel, welcome factor into the test routine, rather than to the beneficial or neutral effects of alcohol per se. Certainly, the poor initial performance of the group, shown in figure 9, was largely due to the gross impairment of subject H (fig. 10) and one other subject in their first test session. This impairment confirms many previous findings of poor initial performance, followed by improvement due to the "warmup" effect, but the gross degree of initial impairment of two subjects suggests that overnight drinking was an additional cause of their extremely poor performance. This suggestion is compatible with the impressions received from cross-questioning within the group.

Some results of the experiments conducted in collaboration with HumRRO scientists are shown in figures 11 and 12, and there is some correlation between signals missed in the HumRRO test of vigilance and the number of long response latencies found in the NRC stress-analyzer experiment. Complete identity of results was not expected, nor found, because subjects were under direct observation and supervision in the stressalyzer test but were isolated in a booth during the HumRRO tests of vigilance and simulated driving. Despite

large differences in the degree of coercion employed, it is probable that increases of long latencies in the stressalyzer test lasting 1 or 2 minutes, were reflecting the lowered attentiveness that increased the number of missed signals in the vigilance test lasting 90 minutes. In the former, unlike the latter test, increased impairment could not be attributed to boredom or to withdrawal from the task.

Figure 1 shows that long latencies on improbable steps usually produce correct responses, and an increased number of long latencies could result from increased stress and a compensatory increase in the effort expended to maintain an adequate standard of overt tracking performance on the stressalyzer test.

This explanation is compatible with data presented in figure 2. At the 28-hour stage, there was a large increase in the percentage of long latencies, although the statistic, reflecting overall tracking ability, did not show a similar abrupt rise until the 36-hour stage. With further sleep deprivation of 44 hours, and then 48 hours, long latencies increased and there was a concurrent deterioration in overall tracking ability as shown by the statistic. Figure 7 illustrates a general group tendency to perform worse by night than by day; the well-known effect of the circadian rhythm. Subject 8, as shown in figure 6, had an abnormally large increase in long latencies about 28 hours after starting the test, but this was associated with slight improvement in overall tracking scores as denoted by the statistic. A large increase in his long latencies and serious impairment of tracking ability occurred about 44 hours after commencing the test. He was the only subject who developed a psychosomatic disability, and who failed to complete the second part of the test wherein sleep was disturbed.

Figure 9 shows little difference in scores in the first and last session of a group, although the subjects were intoxicated and had been deprived of sleep for about 32 hours when the last runs were made. However, there was a high percentage of long latencies in the last session, and in other groups and individuals such increases immediately precede deterioration in tracking performance. Many writers have considered that these so-called blocks or lapses reflect brief periods of sleep, or a condition like sleep. In sharp contrast, the present view is that, at least on their first appearance, these blocks probably reflect periods wherein considerable energy is expended in an effort to maintain adequate tracking performance. The traditional explanation may be valid at a later stage in sleep deprivation, when tracking performance indicates considerable impairment.

Williams et al. (ref. 7) report that in their experiments, using subject-paced tasks, sleep deprivation produced a decrease in speed but little or no change in accuracy. In contrast, both speed and errors increased with sleep loss in the present experiment, using the subject-paced task of mirror tracing. It is possible that the relative effect of sleep loss on speed and error is greatly affected by the degree of continuity of tasks.

Differences in the vulnerability of various tasks to sleep deprivation have been attributed to differences in experimental conditions such as the interest, complexity, and duration of a task, or the presence or absence of monotony or knowledge of results. For some purposes, Broadbent and Williams et al. (refs. 6 and 7) found it useful to distinguish between paced tasks wherein the experimenter controls the onset and timing of a stimulus, as in a vigilance test, and unpaced tasks such as mirror tracing where the subject determines his own speed and timing.

In the NRC and HumRRO studies, there were large differences of impairment in tasks of equal duration when full knowledge of results was given in all cases. The considerable impairment on the stressalyzer cannot be attributed to monotony because the test lasted only 2 minutes. There was severe impairment in two paced tasks, of vigilance and stressalyzer tracking, but comparatively little deterioration on the paced task of simulated driving. The distinction proposed by Wilkinson (refs. 2 and 5) to account for such differences between tasks, is based on the mentalistic qualitative term and concept of relative interest, but in

conflict with this hypothesis, the opinion of all subjects was that the stressalyzer was more interesting than mirror tracing. The distinctions proposed to date are not valid for classifying tasks according to their vulnerability to fatigue.

The real basis for differences in the vulnerability of various task to stress, is the relative amount of external information that must be processed for efficient performance, the continuity of the task, and the amount of advance information on response requirements that is available in the task. No information is conveyed by a statement of the obvious, and more information is conveyed by the occurrence of an improbable than a probable event. Figure 14 shows that stress affects decisions on improbable rather than probable events. The NRC unpaced task of mirror tracing, and the HumRRO paced task of simulated driving, were continuous and gave anticipatory information on the responses required. The tasks did not demand unexpected responses and little external information had to be processed for efficient performance. In sharp contrast, the stressalyzer and the vigilance task required responses to improbable events or to signals that appeared at random, unpredictable intervals; the occurrence of such stimuli presented considerable information.

An important finding reported by Wilkinson (ref. 4) was that knowledge of results reduced the adverse effects of sleep deprivation. In contrast to Wilkinson's interpretation, however, it is concluded by this writer that the major effect of knowledge of results is to enhance objective information, rather than subjective interest, thus compensating for the reduction by stress, of the rate at which a subject can process information.

ACKNOWLEDGMENTS

G. Rowlands of the Royal Aeronautical Establishment, Farnborough, U.K., provided capable assistance on some of the tests, while seconded to the Division of Mechanical Engineering. Mr. R. Leonardo conducted many of the tests and analyses. Dr. L. D. Cannon, E. H. Drucker, and J. R. Ware, all of HumRRO, kindly provided data on the experiments that they conducted on vigilance and simulated driving. D. Gibbs and R. Hughes made valuable contributions to the design of instruments used in the studies. Thanks are also due to members of the Defence Research Board of Canada, and of the Armed Services of Canada and the United States of America for considerable help in arranging and conducting the experiments.

REFERENCES

1. Gibbs, C.B.: Probability Learning in Step-Input Tracking. *Brit. J. Psychol.*, vol. 56, nos. 2 and 3, 1965, pp. 233-242.
2. Gibbs, C.B.: The Effect of Minor Alcohol Stress on Decision Processes in a Step-Tracking Task. *IEEE Trans. Human Factors in Electronics*, vol. HFE-7, no. 4, Dec. 1966, pp. 145-150.
3. Borkenstein, R.F.; Crowther, R.F.; and Ziel, W.B.: The Role of the Drinking Driver in Traffic Accidents. Dept. of Police Administration, Indiana University, 1964.
4. Wilkinson, R.T.: Interaction of Lack of Sleep with Knowledge of Results, Repeated Testing and Individual Differences. *J. Exptl. Psychol.*, vol. 62, no. 4, 1961, pp. 263-271.
5. Wilkinson, R.T.: Sleep Deprivation. *The Physiology of Human Survival*. (Edholm, V.G. and Bacharach, A.L., eds.) Academic Press, London, 1965.
6. Broadbent, D.E.: Noise, Paced Performance and Vigilance Tasks. *Brit. J. Psychol.*, vol. 44, 1953, pp. 295-303.
7. Williams, H.L.; Lubin, A.; and Goodnow, J.J.: Impaired Performance with Acute Sleep Loss. *Psychol. Mono.*, vol. 73, no. 14, 1959.

14. Predictive Compensation in Time-Delay Manual Control Systems

*Eric A. King-Smith
University of Michigan*

The effect of incorporating predictive compensation in a manual control system wherein the controlled plant includes a transport delay is investigated. Controllability is found to be greatly improved and operator effort reduced; the system performance (IAE) is also found to be improved generally within the limits imposed by the signal bandwidth-plant delay time product.

Based on operator transfer function measurements, the crossover model is found to be a reasonable fit also for time-delay plants. Two modified crossover models are suggested that provide progressively closer fit than either the standard or extended (α) crossover models.

TIME-DELAY SYSTEMS AND PREDICTIVE CONTROL

Almost since the birth of automatic control systems, the control of plants that include a pure time delay, or "transport delay," has been a prime problem. The amount, scope, and time spread of the literature on the subject is adequate evidence (refs. 1 to 6).

There are several approaches through which adequate control of these plants may be implemented and the best controller selected in purely technological applications. Among these are the classic three-term controller (refs. 1 to 3), carefully selected data-sampling rates in conjunction with other linear or nonlinear compensation (ref. 7), and so on. Here we are going to consider the controller to be a human operator, in particular one employed in a visual-manual control task; this controller has fairly well-definable, adaptive characteristics if linearized, with bounds on the limits of adaptation of the various parameters determining the characteristics (refs. 11 and 12). With these limitations, supplementary compensation is generally necessary to enable the human operator to deal effectively with plants involving a time delay, the subject of this paper. A prime application example, of course, is the control of a vehicle on the Moon by an operator on Earth which involves a round-trip signal delay of 2.6 seconds (refs. 13 and 14).

One particular group of compensation methods for time-delay systems makes use of another time delay as part of the equalizer for the improvement of system stability and performance. Particular interest will be focused here on the "predictive control" compensation proposed first by O. J. M. Smith in 1950's (refs. 8 to 10).

Figure 1(a) shows the classical linear control system with the plant characteristic shown partitioned to separate off its time-delay characteristics from the remainder of its

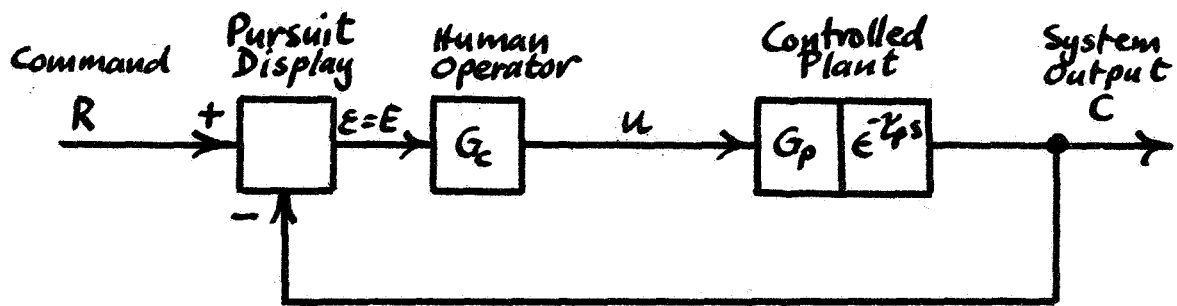


Figure 1a. —Manual control of a plant including a time delay τ_p .

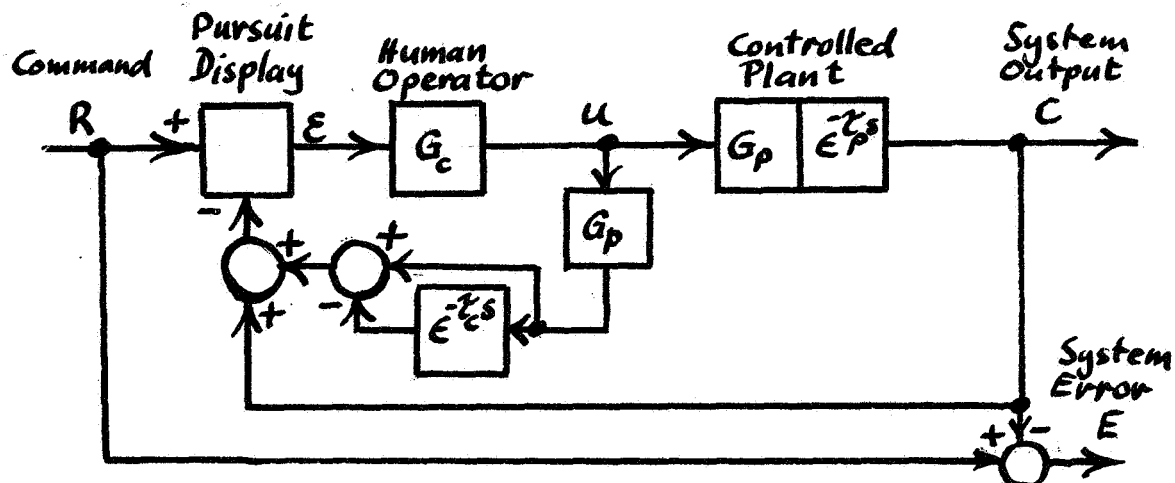


Figure 1b. —Predictive compensation τ_c applied to the system of figure 1a.

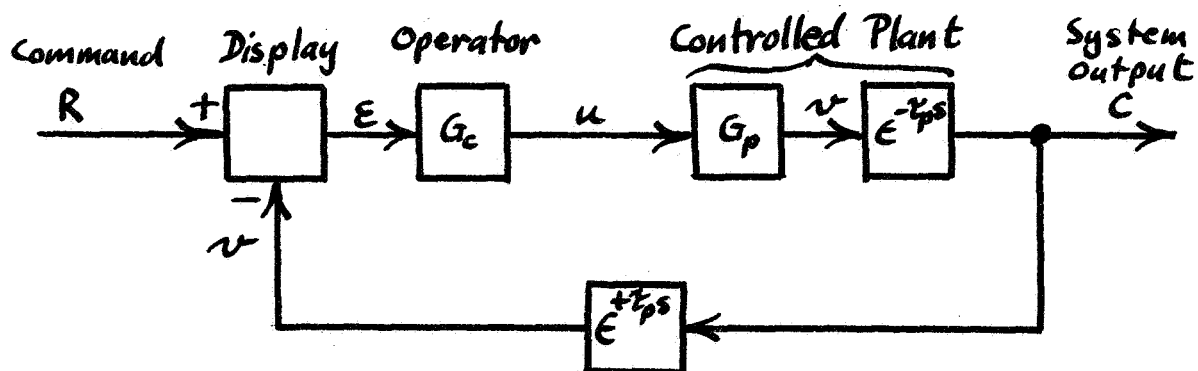


Figure 1c. —Rearrangement of figure 1b with $\tau_c = \tau_p$ to show the effective prediction applied to the signal feedback to the display. Note that the undelayed plant output v appears on the display.

dynamics. It must be assumed, however, that physically these are inseparable within the plant, i. e., no measurement can be made of the variable v . A unity feedback path is shown for simplicity. The open-loop transfer function is

$$G_{OL}(s) = G_c(s) e^{-\tau} p^s G_p(s) \quad (1)$$

and the system overall transfer function

$$T(s) = \frac{G_{OL}}{1 + G_{OL}} = \frac{G_c e^{-\tau} p^s G_p}{1 + G_c e^{-\tau} p^s G_p} \quad (2)$$

with system characteristic equation

$$1 + G_c e^{-\tau} p^s G_p = 0 \quad (3)$$

The delay term $e^{-\tau} p^s$ will ordinarily be a destabilizing factor in any such system, and, if τ_p is comparable to, or greater than the characteristic frequencies of G_p , it will dominate the design for controller characteristics G_c . In terms of a manual control task, this means that system delay will almost inevitably make the operator's job more difficult, and, with the simple system of figure 1, could make it impossible for him to hold stable.

Consider now figure 1(b) which is the original figure 1 with a subsidiary feedback loop added. It is noted that this loop includes a copy (model) of the plant linear transfer function G_p and a delay of value τ_c . Also to be noted is that the input to the controller is no longer the system error $E \triangleq R - C$, but a variable ϵ which is, to a human operator, the deviation that he would see on a compensatory type display, or the difference between command R and some other displayed variable that he is asked to track to R as closely as possible.

If the open-loop transfer function is now taken from the point of view of the controller-operator, we obtain

$$G_{OL} = \{e^{-\tau} p^s G_p + G_p [1 - e^{-\tau} c^s]\} G_c \quad (4)$$

$$G_{OL} = G_p G_c \{1 + (e^{-\tau} p^s - e^{-\tau} c^s)\} \quad (5)$$

If τ_c , now designated the "compensation delay," is made equal to τ_p , we will obtain simply

$$G_{OL} = G_p G_c \quad (6)$$

meaning that the operator sees merely the plant linear dynamics G_p without the delay τ_p . Almost inevitably this will make the operator's task easier, in fact it may well change a completely uncontrollable plant into a quite docile one, since the characteristic equation is now simply

$$1 + G_c G_p = 0 \quad (7)$$

A block-diagram manipulation of figure 1(b) with $\tau_c = \tau_p$ yields figure 1(c), where a predictive element is now explicit, the existence of which reinforces the title assigned to the compensation technique.

The immediately evident disadvantage is that the operator is no longer conscious of the true system error E , but only of the substituted variable ϵ . Any normal performance measure will yield a higher index of performance (or lower error score) when operating on ϵ than on E , as shown below. Whether the performance index based on E is improved depends on several factors, principally the relative stabilities of the system with and without predictive compensation, and the spectral characteristics of the command signal in relation to the plant time delay τ_p .

The compensation delay τ_c may not necessarily be set exactly equal to τ_p , and τ_c thus becomes an assignable parameter in the design of the system. If $\tau_c < \tau_p$ the system is said to be undercompensated; if $\tau_c > \tau_p$ the system is overcompensated.

THE BASIC SYSTEM ERROR DUE TO SIGNAL DELAY THROUGH THE PLANT

For the uncompensated original system (fig. 1(a)), the closed-loop system transfer functions are

$$T_0 = \frac{G_c e^{-\tau} p^S G_p}{1 + G_c e^{-\tau} p^S G_p} \quad (8)$$

and for the fully compensated system

$$T_1 = \frac{G_c e^{-\tau} p^S G_p}{1 + G_c G_p} \quad (9)$$

so that the system error is

$$E_1 = R \left[1 - \frac{G_c e^{-\tau} p^S G_p}{1 + G_c G_p} \right] \quad (10)$$

It is not feasible for the operator to track the command perfectly in the uncompensated case, but could be postulated in the fully compensated case, e.g., for a very narrow bandwidth signal. This would make $\epsilon = 0$, and represent the ultimate performance as the operator could see it. Such a condition implies

$$\frac{G_c G_p}{1 + G_c G_p} = 1 \quad (11)$$

so that

$$E_1 = R \left[1 - e^{-\tau} p^S \right] \quad (12)$$

or, that the system error is equal to the difference between the command signal and itself delayed by τ_p .

Consider now that R is band-limited gaussian white noise with autocorrelation function $\phi_{RR}(\tau)$. Then the mean square value of the error between $R(t)$ and $R(t-\tau_p)$ is

$$\bar{E}^2 = 2[\phi_{RR}(0) - \phi_{RR}(\tau_p)] \quad (13)$$

This may be interpreted simply to mean first, that a zero plant delay with perfect tracking will yield zero \bar{E}^2 , as expected; second, if the delay τ_p is much greater than the reciprocal bandwidth, resulting in $\phi_{RR}(\tau) \rightarrow 0$, $\bar{E}^2 \rightarrow 2\phi_{RR}(0) = 2\bar{R}^2$, again an expected result since input and output are then mutually uncorrelated.

In the latter case, the operator would do better to hold the system output at zero (if he could) since then $\bar{E}^2 = \bar{R}^2$ and not $2\bar{R}^2$. However, this strategy is not possible with either pursuit or compensatory displays since the actual system output is not apparent to the operator.

At some intermediate point, perfect tracking will result in a mean square error equal to \bar{R}^2 , which is a good definition of the absolute limit of usefulness of the compensation technique. This condition clearly occurs when $\phi_{RR}(\tau_p) = 1/2$.

For a second-order aperiodic filter shaping the white noise to produce the command signal, as used in the experimental trials, this can be shown to occur when

$$\omega_c \tau_p = 1.68 \text{ or } \omega_c = 1.68/\tau_p \quad (14)$$

where ω_c = corner frequency (filter attenuation = 6 decibels). The actual autocorrelation function is

$$\phi_{RR}(\tau) = e^{-\omega_c \tau} + \omega_c \tau e^{-\omega_c \tau} \quad (15)$$

INITIAL CONCEPTS OF POSSIBLE RESULTS OF PREDICTIVE COMPENSATION IN A MANUAL CONTROL SYSTEM

Let us consider a system incorporating an actual plant delay τ_{pa} , which, without compensation, is controllable by the operator, i.e., he can hold it stable and track the command signal, albeit with considerable error. Suppose that his mean square display error is $[\bar{e}^2]$.

As the compensation is applied, the operator's performance, as measured by mean square error on the display, will generally improve to some minimum near, but not necessarily at, perfect compensation. Figure 2(a) shows this typically; the minimum mean square display error is here denoted $[e^2]_{\min}$, while that occurring for perfect compensation ($\tau_c/\tau_{pa} = 1$) is denoted $[\bar{e}^2]_p$.

The actual form of the curve in figure 2(a) is characteristic of the operator in that it is simply a description of his improvement in performance as the system becomes easier for him to handle. This characteristic is of considerable significance in that not only do the lower error scores with compensation form a measure of the ease of system controllability, but also they provide a measure of the mental effort required of the operator. An improved performance together with a lower mental effort is a desirable goal, but there is always in the background the question of whether optimum overall system performance, when the actual error is unknown to the operator, can, in fact, coincide with minimum operator load.

Consider now a different aspect of the same system. Let it be fully compensated at all times, and let us assume for the moment that the human operator is a perfect tracker. The minimum mean square system error will then be zero at $\tau_p = 0$. According to the derivation in the preceding section, the mean square system error will increase with $\tau_p \neq 0$ until it reaches an asymptotic maximum

$$[\bar{E}^2]_{\max} = 2\phi_{RR}(0) = 2\bar{R}^2 \quad (16)$$

Figure 2(b) is typical of operation with a second-order band-limited white noise command signal; a typical actual plant delay τ_{pa} is also indicated, together with its corresponding value of $[\bar{E}^2]_a$.

To a first approximation the curves of figure 2(a) and (b) may be considered as complementary, i.e., increasing τ_c from zero to some actual plant delay τ_{pa} leads the operator to believe that the system delay is progressively reduced to zero; more correctly, the apparent delay between his manual action and the display response seems to decrease from τ_p to zero. (A limited bandwidth is clearly necessary for this to be true; a rule-of-thumb specification would be that implied by equation (14), i.e.,

$$\omega_b \tau_p \approx 1.68 \quad (17)$$

where ω_b is the bandwidth of the signal allowed through to the display.)

Combining the two characteristics to allow for the operator's actual tracking performance, it is seen that a net improvement in performance will occur for perfect predictive compensation if

$$[\bar{E}^2]_o - [\bar{E}^2]_p > [\bar{E}^2]_a \quad (18)$$

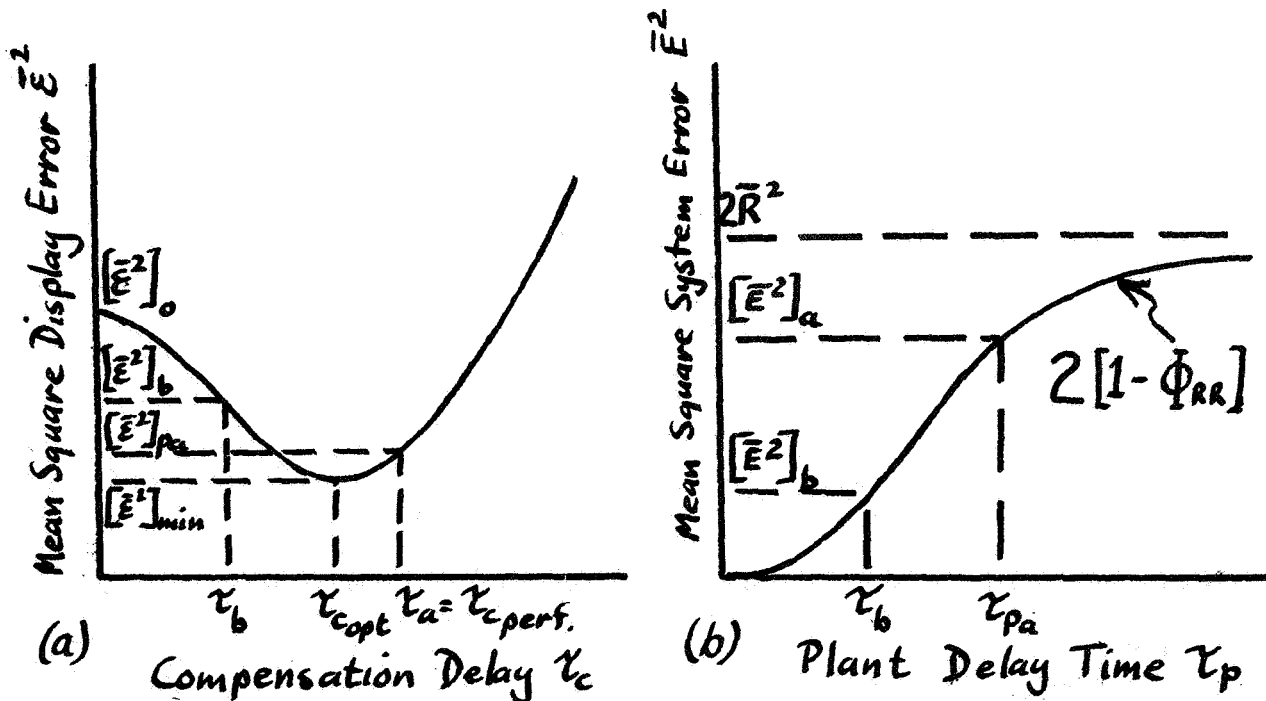


Figure 2. —(a) Typical variation of mean square display error with compensation delay.

(b) Mean square system error as a function of system delay for perfect tracking.

when the improvement in system mean square error will be (approximately)

$$[\bar{E}^2_{imp}]_p \approx [\bar{\epsilon}^2]_o - [\bar{\epsilon}^2]_p - [\bar{E}^2]_a \quad (19)$$

For other than perfect compensation, the improvement will be given by

$$[\bar{E}^2_{imp}]_b \approx [\bar{\epsilon}^2]_o - [\bar{\epsilon}^2]_b - [\bar{E}^2]_b \quad (20)$$

where b is relevant to the nonperfect compensating delay τ_b .

What value of τ_b will bring about the minimum value of $[\bar{E}^2_{imp}]$ depends on the relationship between the two curves. Clearly some immediate improvement should be noted for small amounts of compensation if the characteristics are typical as shown, but that is about all that can be predicted at this stage. As a guide to further prediction, however, we may note that a narrower bandwidth signal will lower the curve in figure 2(b) throughout, generally enhancing the ultimate improvement possible, and vice versa.

EXPERIMENTAL TRIALS

A series of trials were run with three male subjects, aged 21, none of whom had had manual control task experience previously. An analog computer was used to simulate the plant linear dynamics and the compensating loop delay (by Padé approximation) and dynamics; it was also utilized to compute the integrated absolute error scores for both system error and display error over its programed measuring period of 2 minutes for each trial.

A tape recorder was used to generate the plant delay, this being variable in steps according to the recorder tape speed; delays of 180 ms, 360 ms, 720 ms, and 1440 ms were thus available. For long delays an additional 400 ms delay (Padé approximation) was added to the latter two to provide 1120 ms and 1840 ms delays, respectively.

The display on a 5-inch oscilloscope was of the pursuit type, the command being a vertical line approximately 2 inches long and the tracking indicator a spot that bisected the line when on target. Deflection was limited to the horizontal direction and between bounds of ± 6 centimeters from the center of the screen. A force stick was used for control by the subject, the sensitivity being 0.5 V/lb.f. Intercommunication was provided between the trial supervisor's position at the computer, and the subject in the isolation booth; when not actually communicating, a moderate level white noise was fed to the subject's earphones to mask any external noises which might have distracted him.

The signal to be tracked was a pseudorandom telegraph signal of clock rate 1.8 per second, bandlimited to 0.5 rad/sec by a critically damped second-order filter (ref. 15).

A general view of the setup is shown in figure 3(a) with the isolation booth at the left; a closeup of the subject's position in the booth appears in figure 3(b). The distance between the subject's eyes and the oscilloscope screen was 26 to 30 inches, depending on the subject's adopted attitude.

Three different system dynamics were used:

proportional: transfer function $K=1$

single-integration (velocity): transfer function $K/s = 2/s$

double-integration (acceleration): transfer function $K/s^2 = 2/s^2$

With these characteristics, the following obtained with the three different plant dynamics:

proportional: 0.5 lb.f \rightarrow 6 cm deflection on CRO

single-integration: 0.5 lb.f \rightarrow 12 cm/sec deflection on CRO

double-integration: 0.5 lb.f \rightarrow 12 cm/sec² deflection on CRO

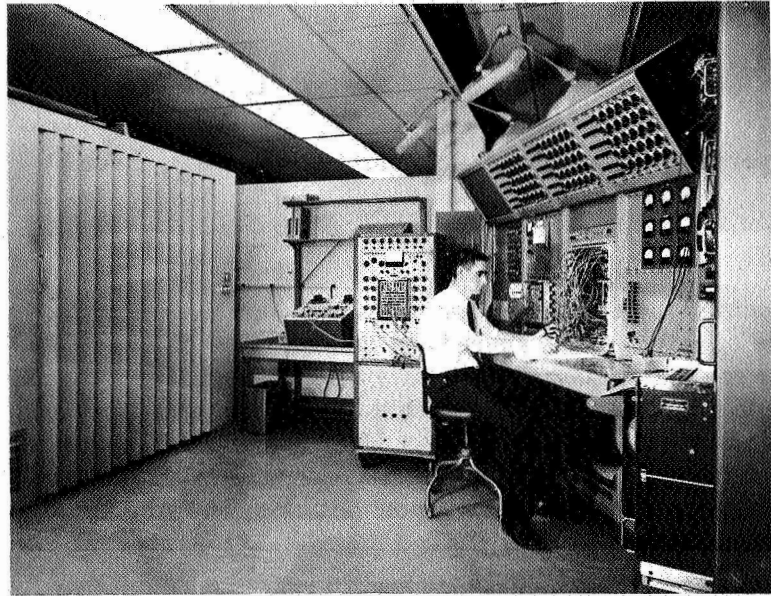


Figure 3a.—General view of the experimental setup; isolation booth is at the left.



Figure 3b.—Operator engaged in the tracking task.

Each subject was given 7 days training prior to the evaluated trials. Each day's trial lasted approximately 1 hour within which period a 2-minute run was made on each type of system - K, K/s, K/s² - with delays 0 ms, 180 ms, 360 ms, 1440 ms. At the end of 7 days, all subjects appeared to have asymptoted to constant performance on all systems.

Each trial for serious evaluation lasted a total of 2 1/4 to 2 1/2 minutes, of which 2 minutes was actually utilized for computation of the integral absolute errors of display and system. The initial 15 to 30 seconds was used to allow the subject to settle down into the task. Short delays (0, 180, and 360 ms) were run as a separate group of trials to long

delays (1120 and 1840 ms); within each day's set of trials, a run was made on each type of system with zero compensation, and τ_r at -100, 0, +100, +200 ms for short delays, and -200, -100, 0, +100, +200 ms for long delays.

Tests were also run during the long trials of the operator's ability to hold the basic system (i.e., no delay or compensation) on a center point; such measurements provided the measure of the subject's basic noise level in dealing with each type of system.

The tape recorder used to generate the plant delay also recorded system output, system error, compensating signal and force-stick output for later spectral analysis processing.

ERROR SCORE RESULTS AND COMMENTS

The normalized integrated absolute value of the command signal, on the same basis as that measured for the error scores, was 1.00, i.e., if the system output were held perfectly stationary, an error score of unity would result.

If perfect tracking were performed on the signal delayed many times its reciprocal bandwidth, the error score would be expected to be $\sqrt{2} = 1.41$.

ERROR SCORE RESULTS

Figure 4 shows the IAE display error scores averaged over the three subjects for each of the three types of system as a function of relative compensation delay. Apart from the differences in actual magnitude of the error scores, there is a significant difference in the value of relative compensation at which the error score is reduced to a minimum, as follows:

Type of system	τ_r for IAE_{min}
K	-35 ms approx
K/s	+30
K/s ²	+170

The trend indicated is supportable by considering the actual effect of nonperfect compensation on the open-loop transfer function. This is

$$G_L = G_c [1 - e^{-\tau_c s} + e^{-\tau_p s}] G_p \quad (21)$$

If the signal bandwidth is relatively small, we can make the approximating substitutions $e^{-\tau_c s} \approx 1 - \tau_c s$, $e^{-\tau_p s} \approx 1 - \tau_p s$ to obtain

$$G_L \approx G_c G_p [1 + (\tau_c - \tau_p) s] = G_c G_p [1 + \tau_r s] \quad (22)$$

Hence, if $\tau_r > 0$, the effect of the compensation as it appears to the operator is akin to "quickening" (added velocity feedback); an added lag on top of the G_p dynamics is the apparent effect if $\tau_r < 0$.

On the basis of the crossover model for the human operator (ref. 11) the result above is to be expected in that a second-order system would require a derivative function for compensation, and the proportional system a lag function to enable the operator to behave in his preferred mode of a straight amplifier.

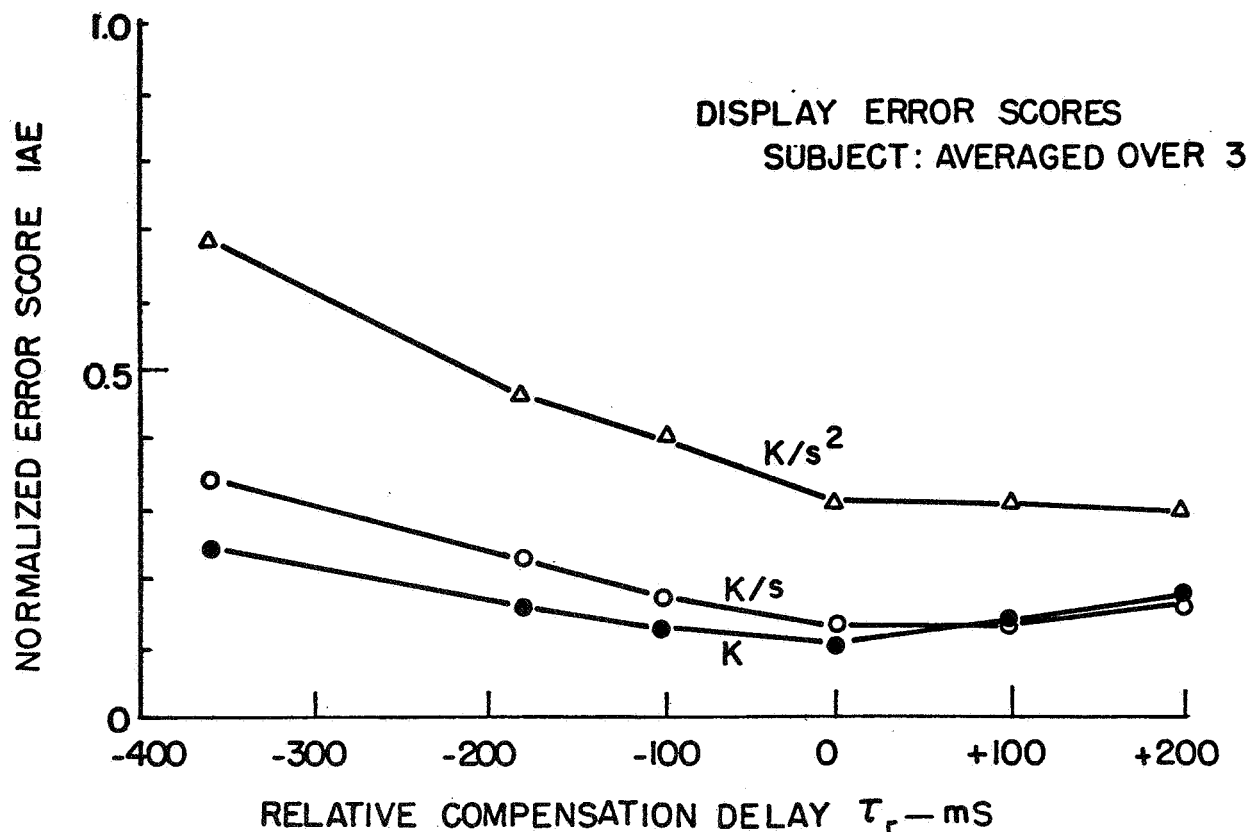


Figure 4. —Display error scores as a function of relative compensation delay.

Figures 5(a), (b), and (c) display the system error scores for the three types of system for short delays (up to 360 ms). Display error values are also given here for comparison, revealing the error score component due to the plant delay.

At the side of the diagram is a stepped reference. The right-hand level of the step indicates the error score which would result from a perfectly tracking operator; the left-hand level includes the basic noise of the operator evaluated from the center-holding tests.

It is seen that for a proportional system compensation of the type used here is of no use in improving performance per se; the operator does best without it. However, as the display error scores indicate, and supported by the subject's response to providing compensation, the mental workload is considerably reduced; in other words, the workload is shifted from the operator to the compensation system, but with a tendency toward a net degradation in overall system performance.

For type 1 (K/s , single-integration) systems, a small, but significant, improvement in system performance is obtained, combined with a greater reduction in operator load than with the proportional system. For type 2 (K/s^2 , double-integration) systems, a very useful improvement in system performance is obtained, and a relatively great reduction in workload (due to the low stability of the system without compensation) is achieved. Note that overcompensation applied to a zero-delay plant also produces an improvement; this is tantamount to quickening applied to an acceleration system, as pointed out earlier.

Figure 5 applied to bandwidth delay values of 0, 0.09, and 0.18, for which very effective performance improvement was predicted as being possible. Figure 6 for a single-integration system shows the error score results for plant delays of 1120 ms and 1840 ms,

for which the corresponding bandwidth-delay figures are 0.56 and 0.92; a relatively marginal improvement only is indicated, particularly for the 1840 ms situation.

The results show that system error scores are not significantly affected by applying compensation, although there is a relatively enormous improvement in display error score and reduction in operator load. In the case of the 1840 ms uncompensated system, the operators could barely hold the system stable for the whole of the 2-minute trial period, and it is, at first sight, a little surprising that a higher uncompensated error score was not obtained.

These results appear to show an almost perfect tradeoff between reduction of task difficulty, and the amount of compensation provided by the operator to reduce system error. When subjects were conscious of the long delay, it was observed that they used their knowledge of the statistics of the signal to be tracked to reduce the error score, e.g., if the command moved almost to maximum excursion of the display, the subject knew that it almost invariably returned to more moderate deflections rather soon, and did not attempt to follow

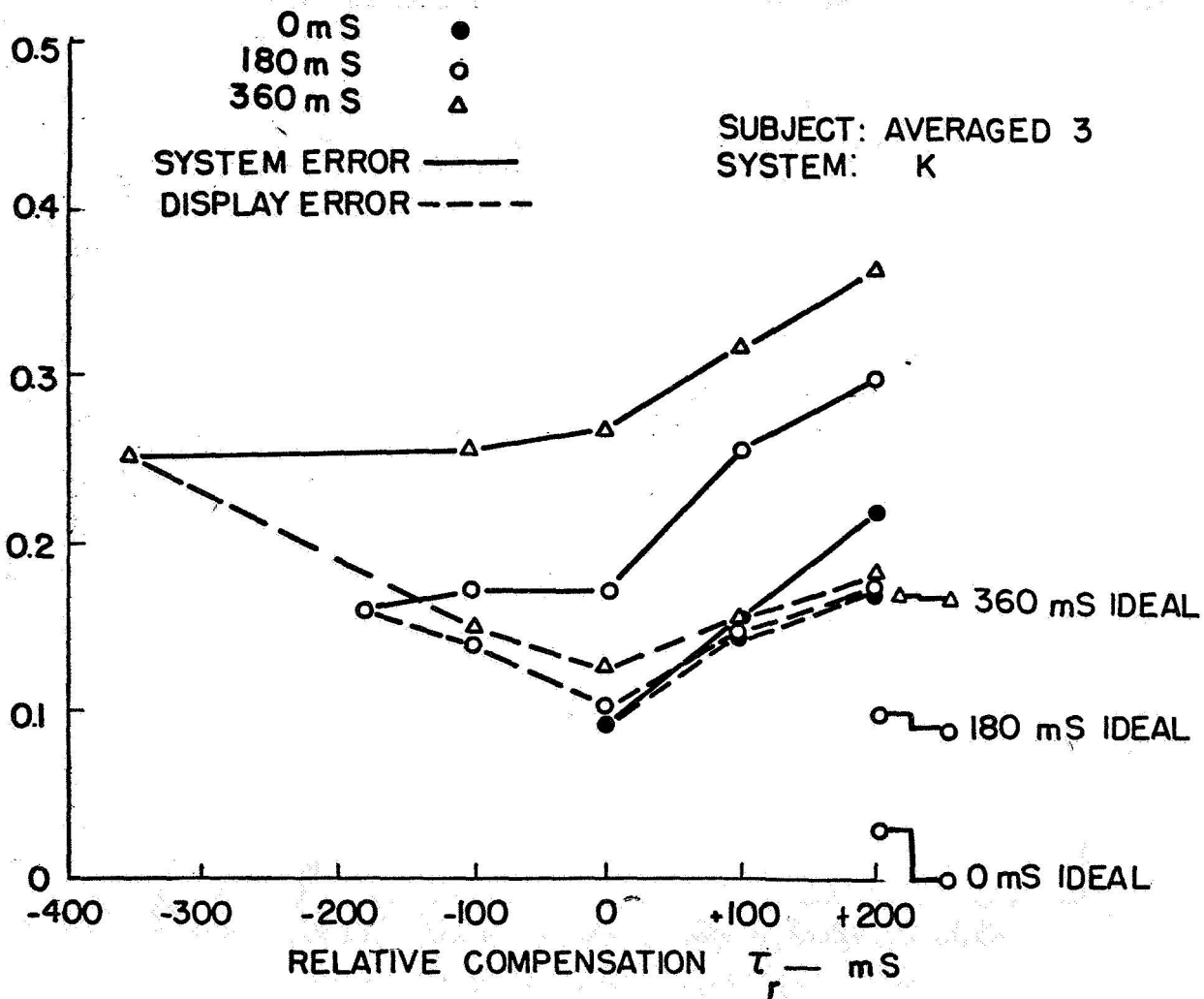


Figure 5a. —System and delay error scores for plant delays of 0, 180, and 360 msec for type 0 ("proportional") system.

fully, but to catch it on return; a lower error score results from this strategy than if an identical technique were used over the entire display range. This observation seems to explain adequately why error scores do not improve between zero compensation and $\tau_r = -200$ ms.

Error score characteristics for proportional and double-integration systems were very similar to those of the single-integration system of figure 6.

SPECTRAL ANALYSIS—CROSSOVER MODEL VALIDATION

To ascertain whether McRuer's simple (or classic) crossover model (ref. 12) of the human operator could be fitted to the control of plants, the tape-recorded data from one subject on a type 1 system was analyzed to yield both open-loop transfer functions and operator transfer functions for all short delay trials, runs under identical conditions being averaged. An example of the order of fit obtained is shown in figure 7. It was concluded that the crossover model could be validly invoked in the analysis and design of such systems.

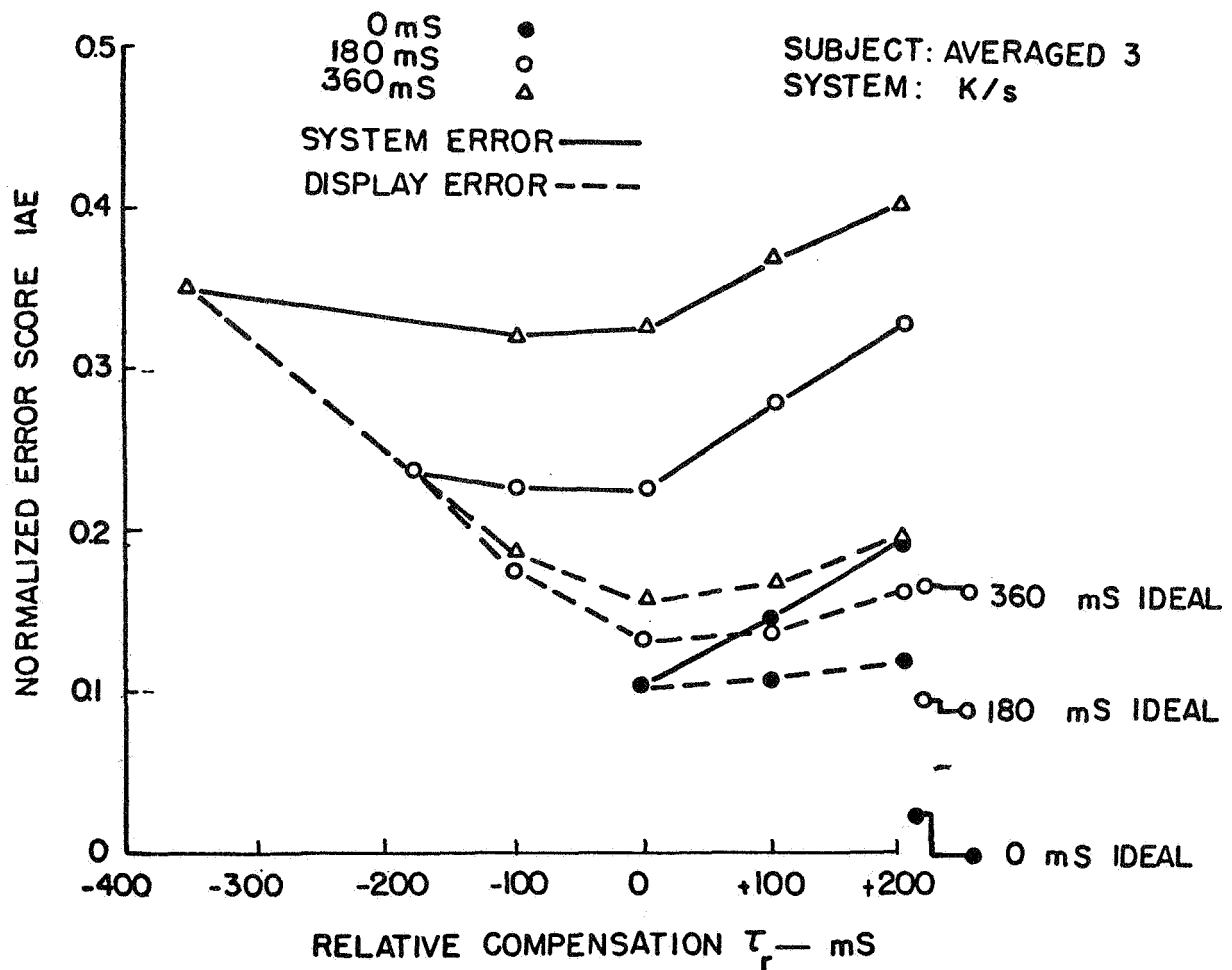


Figure 5b.—System and display error scores for plant delays of 0, 180, and 360 msec for type 1 ("velocity" or "single-integration") system.

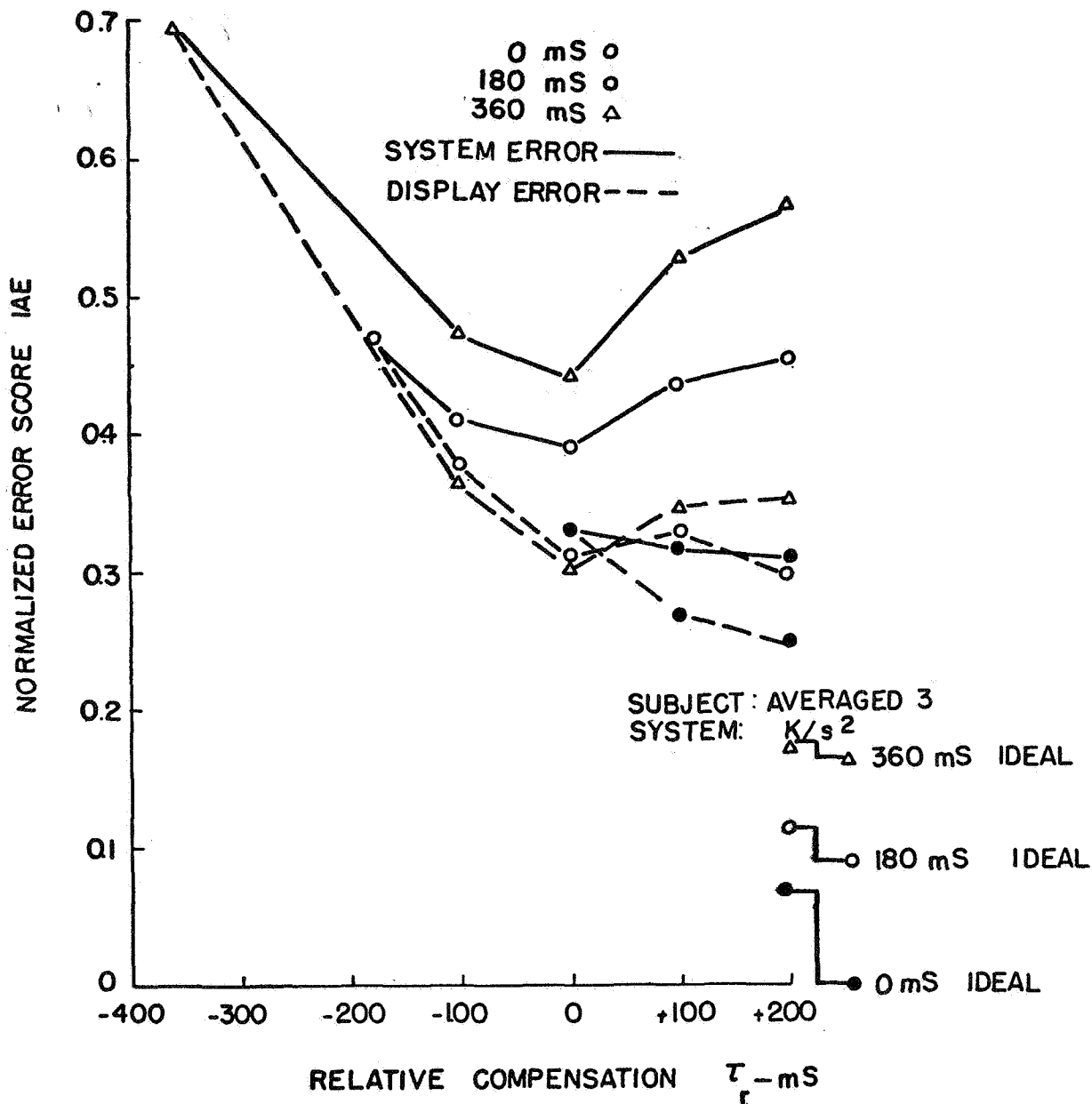


Figure 5c.—System and display error scores for plant delays of 0, 180, and 360 msec for type 2 ("acceleration" or "double-integration") system.

Rather than display all relevant spectral plots, figure 8 shows the variation in cross-over frequency, loop gain, and loop delay as a function of plant delay, with no compensation. The trends are fairly self-explanatory.

Another interesting characterization is the variation of crossover model parameters with relative compensation. This is shown in figure 9. The trend to optimum values - maximum loop gain, minimum loop delay, and maximum crossover frequency - at perfect

compensation, as should be expected for this type of system, is plain, in spite of the apparent plant (to the operator) having the somewhat queer transfer function

$$G_{pa} = G_p [1 - e^{-\tau_c s} + e^{-\tau_p s}] \quad (23)$$

The fit would, in fact, be considered reasonably good in terms of past exercises of this sort, not only in the crossover region, but also right through to much lower and much higher frequencies.

PROPOSALS FOR MODIFIED CROSSOVER MODELS

During the crossover model curve fitting process, the author was struck, as were others, with the fairly consistent low-frequency phase drop in many of the spectral plots, figure 7 being typical. The "extended" crossover model, or α -crossover model

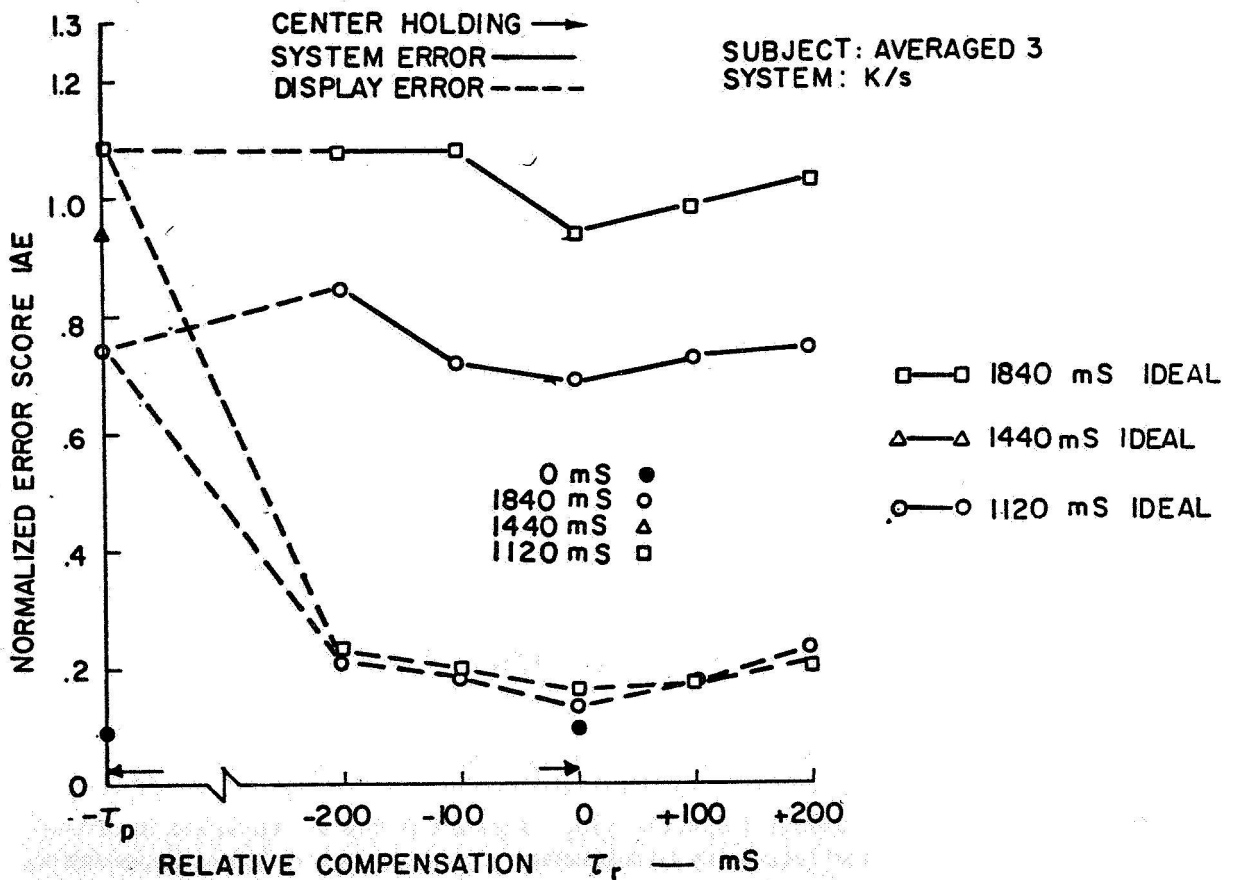


Figure 6.—System and display error scores for plant delays of 1120 msec and 1840 msec.

(ref. 16), was one attempt to obtain a better phase fit at low frequencies. The crossover model assumes a linearized loop transfer function

$$G_{OL}(j\omega) = \frac{K e^{-j[(\tau_c + \tau_p)\omega]}}{j\omega} \quad (24)$$

The α -model modifies this to

$$G_{OLx}(j\omega) = \frac{K e^{-j[(\tau_c + \tau_p)\omega + \alpha/\omega]}}{j\omega} \quad (25)$$

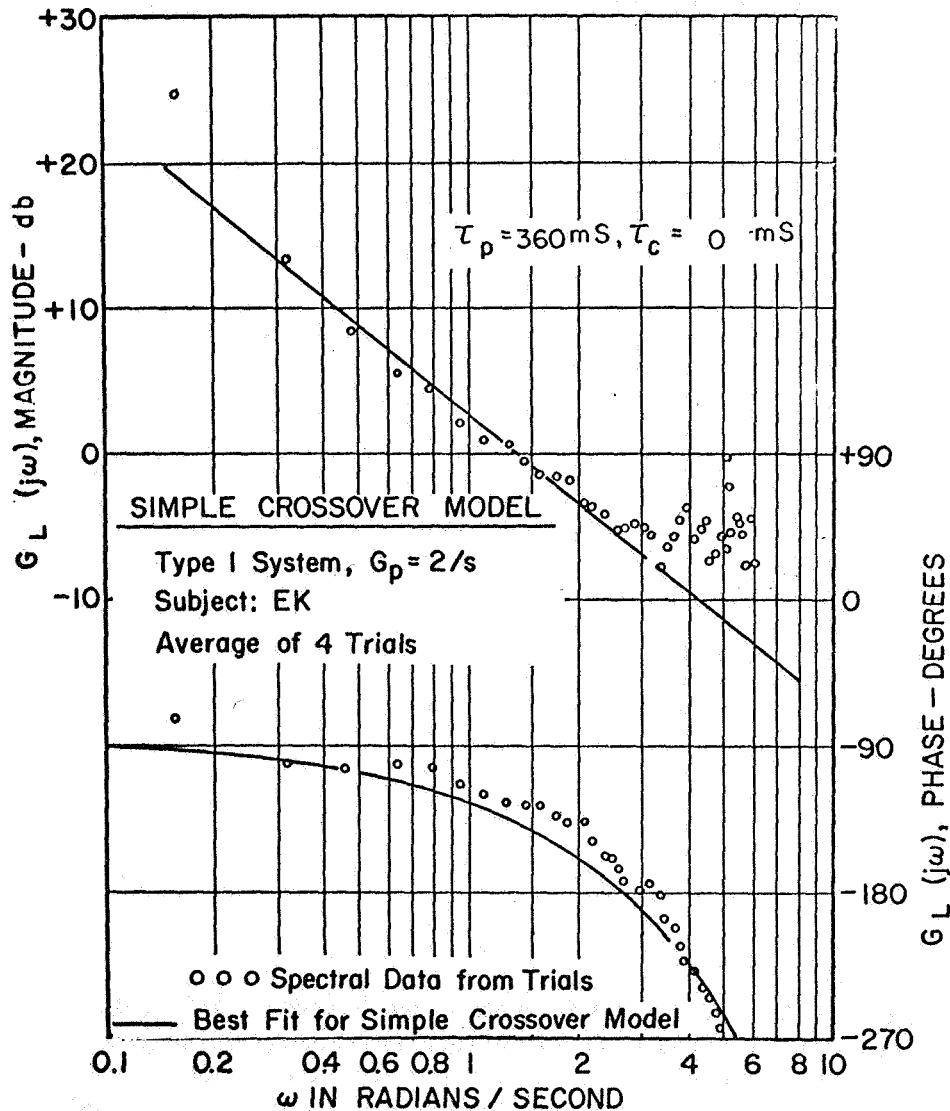
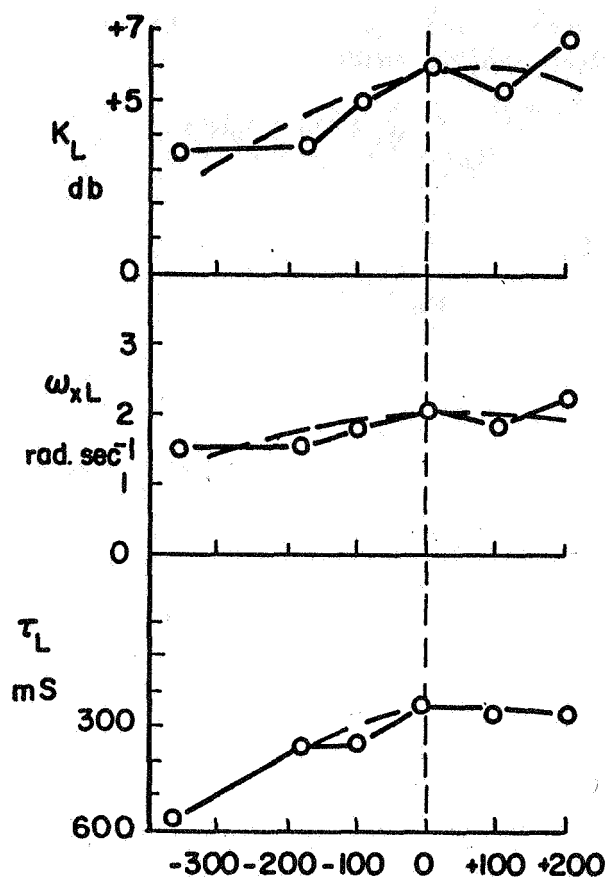


Figure 7. — Example of the simple crossover model spectral data fit in a manual control system with plant time delay.



VARIATION OF LOOP TRANSFER FUNCTION PARAMETERS

MODEL: SIMPLE CROSSOVER

SYSTEM: K/s

SUBJECT: EK

NORMALIZED AND AVERAGED VALUES

Figure 8.—Simple crossover model parameter variations as a function of relative compensation delay.

The purpose of the added parameter α has no physical or physiological meaning; it simply provides means to obtaining a better phase fit at low frequencies. In this sense, it expresses mathematically some peculiar aspect of the human operator's data processing in handling manual control systems.

In the author's view, the main disadvantage of the extended crossover model is that it yields an infinitely negative phase shift as $\omega \rightarrow 0$ or a constant positive phase slope as $\omega \rightarrow 0$. A finite negative phase shift at very low frequencies would imply infinite delay of these signals; an asymptotically positive phase slope would imply an infinite delay of yet higher order.

It should be pointed out here that the crossover model and the extended crossover model have been developed primarily on the basis of pursuit displays in manual control systems. What follows is therefore primarily applicable to pursuit displays, whose low frequency phase lag characteristics do differ significantly from those for compensatory displays (ref. 17).

A study of the loop transfer functions from these trials, and of those from other investigations indicated that a better low-frequency phase fit for pursuit displays would be to a constant phase lag rather than to a constant slope; this conclusion applied particularly well to type 1 systems. Two variations of the crossover model will now be proposed for pursuit displays and illustrated with typical results of the time-delay system trials.

THE MODIFIED CROSSOVER MODEL.—This model differs from the simple crossover model only in that the slope of the magnitude characteristic of the loop transfer function is allowed to vary from its set value of -20 dB/decade. This leads to a three-parameter model for the loop transfer function:

$$T_{Lm} = \frac{K e^{-\tau_e s}}{s^n}$$

the three parameters being loop gain K , equivalent loop delay τ_e , and magnitude characteristic slope $20n$ dB/decade. n also sets the low-frequency asymptotic phase slope at $-90n$ degrees.

Figure 9 also shows this model fitted as solid lines to data taken from a perfectly compensated system, where the model parameters here are $K=+8.3$, $\tau_e = 200$ ms, and $n=1.12$. In this perfectly compensated system, τ_e will be simply the operator's delay.

The use of a nonunity n in this model is no more extreme than using α in the extended crossover model. It has purpose only to improve the fit to the data and cannot be identified physically. Any human operator model contains parameters descriptive of the data processing attributes of the operator as well as other physically identifiable characteristics. The non-unity n simply states that the operator does not process the data to produce an ideal integration within the loop; rather does he select some characteristic between a single and double integration.

Figure 10 illustrates the responses that would result from a constant command to a controller that was characterized by a delay τ_e and various values of n in the model of equation (2). (Such responses are purely hypothetical in that the operator in practice would quickly sense something amiss if the display error remained constant in spite of his corrective actions.)

Characteristic slopes that are not integral multiples of 20 dB/decade have been found in a different context. Pew, Duffendack, and Fensch have reported human operator noise power spectra having slopes over a very wide frequency range of -10 dB/decade (ref. 18); the suggestion is that a similar data processing by the operator is involved also there, as here.

There appear to be no grounds to believe that human data processing in a manual control task is conducted in terms of lumped linear systems. Rather does it seem that the processing is done according to entirely different operator concepts, the result of which are spectral characteristics often having slopes other than multiples of -20 dB/decade and corresponding phases.

One consistent trend was discovered: As the task becomes more difficult, the operator describing function for pursuit displays becomes ever closer to the simple crossover model across the entire measurable spectrum. Figure 7, for a system much more demanding than in figure 9, illustrates this well. Figure 11 illustrates the trend of the model parameters as a function of relative compensation delay for single-integration plants.

THE DOUBLE-SLOPE CROSSOVER MODEL.—Examination of describing function characteristics such as figure 9 showed that even better fit could be obtained with different

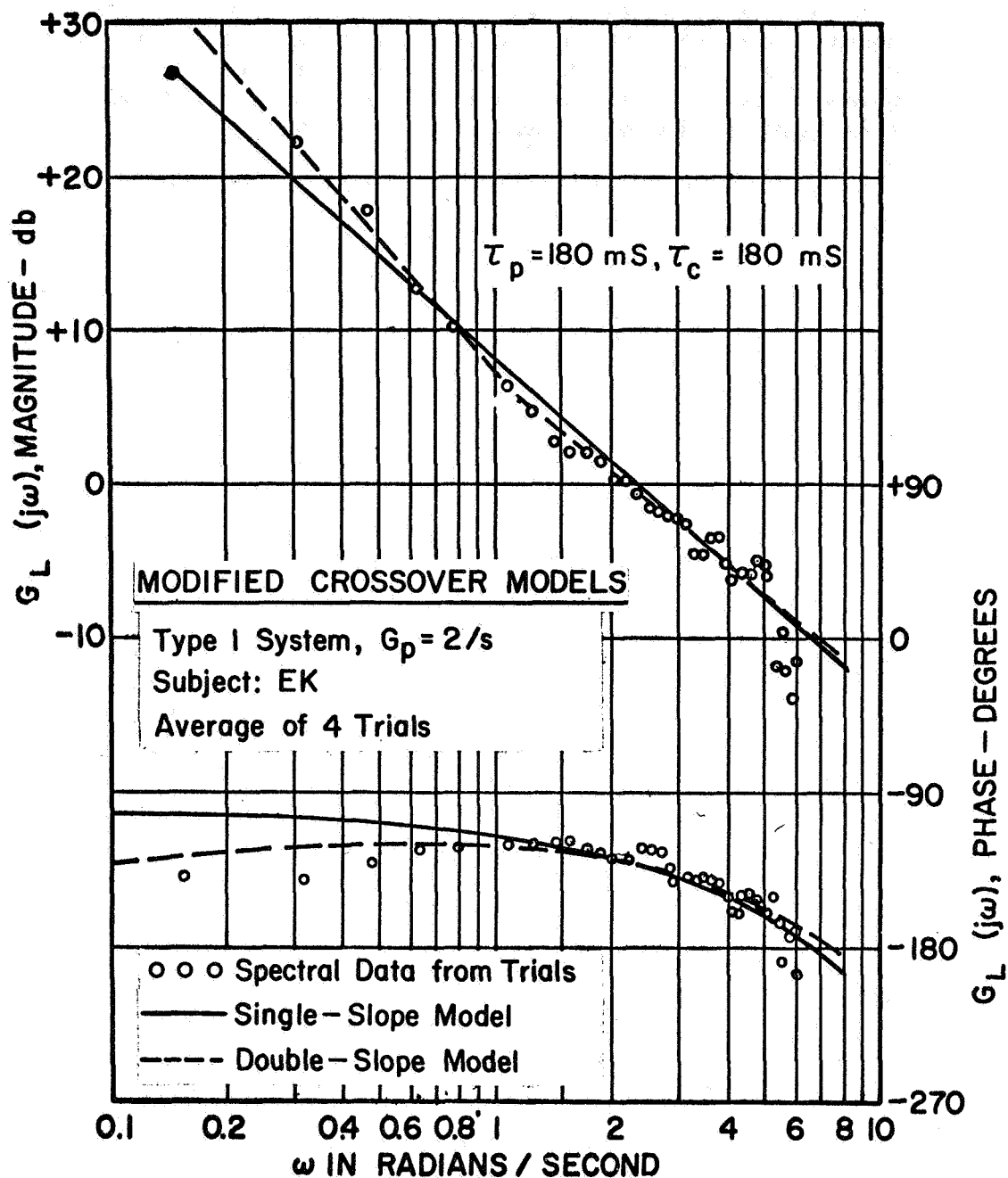


Figure 9. —Example of modified and double-slope crossover models fitted to experimental data.

slopes at high and low frequencies. This leads to a four-parameter model, nominated the "double-slope" crossover model—gain K , low-frequency slope $S \neq 20 \text{ dB/decade}$, break frequency ω_c , and effective time delay τ_e .

Whether such additional refinement is justified is uncertain, except that it allows excellent data fits to the majority of experimental data. In figure 9, the broken lines show the fit of this model to the experimental data for a perfectly compensated system.

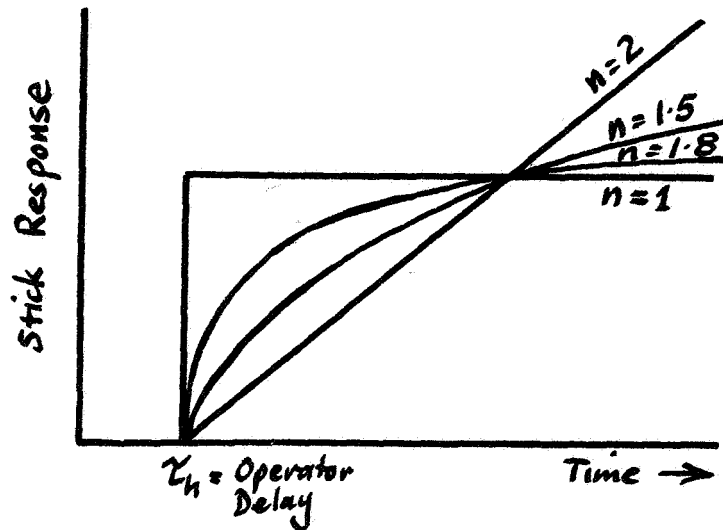
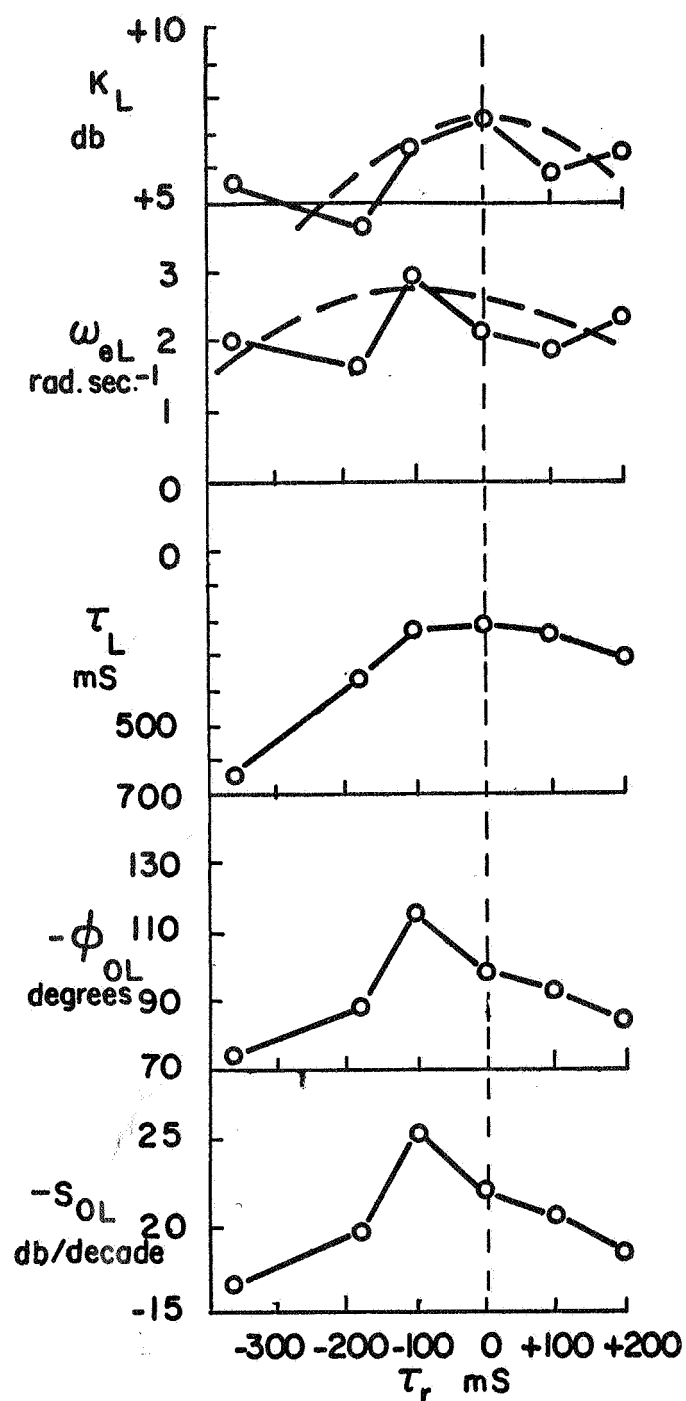


Figure 10.—Hypothetical operator stick responses to a step error signal for various values of n in the modified crossover model.

Figure 12 shows the variation of the model parameters as the relative compensation delay is varied for the predictive compensation schedule.

CONCLUSIONS

- (1) The predictive compensation method is very effective in reducing operator load in manual control systems in which the controlled plant contains considerable pure time delay, but
- (2) It can improve system performance (here measured on an integral absolute error basis) if within the delay time-signal bandwidth constraint, theoretically $\omega_s \tau_d \leq 1.68$.
- (3) The simple crossover model fits the human operator performance well for delays up to 360 ms, using a pursuit display.
- (4) Two variations on the simple crossover model, the three-parameter modified and four-parameter double-slope crossover models, allow even better data fitting than the simple two-parameter crossover model for pursuit displays.



VARIATION OF LOOP TRANSFER
FUNCTION PARAMETERS
MODEL: MODIFIED CROSSOVER
SYSTEM: K/s
SUBJECT: EK
NORMALIZED AND AVERAGED VALUES

Figure 11. —Modified crossover model parameter variations as a function of relative compensation delay.

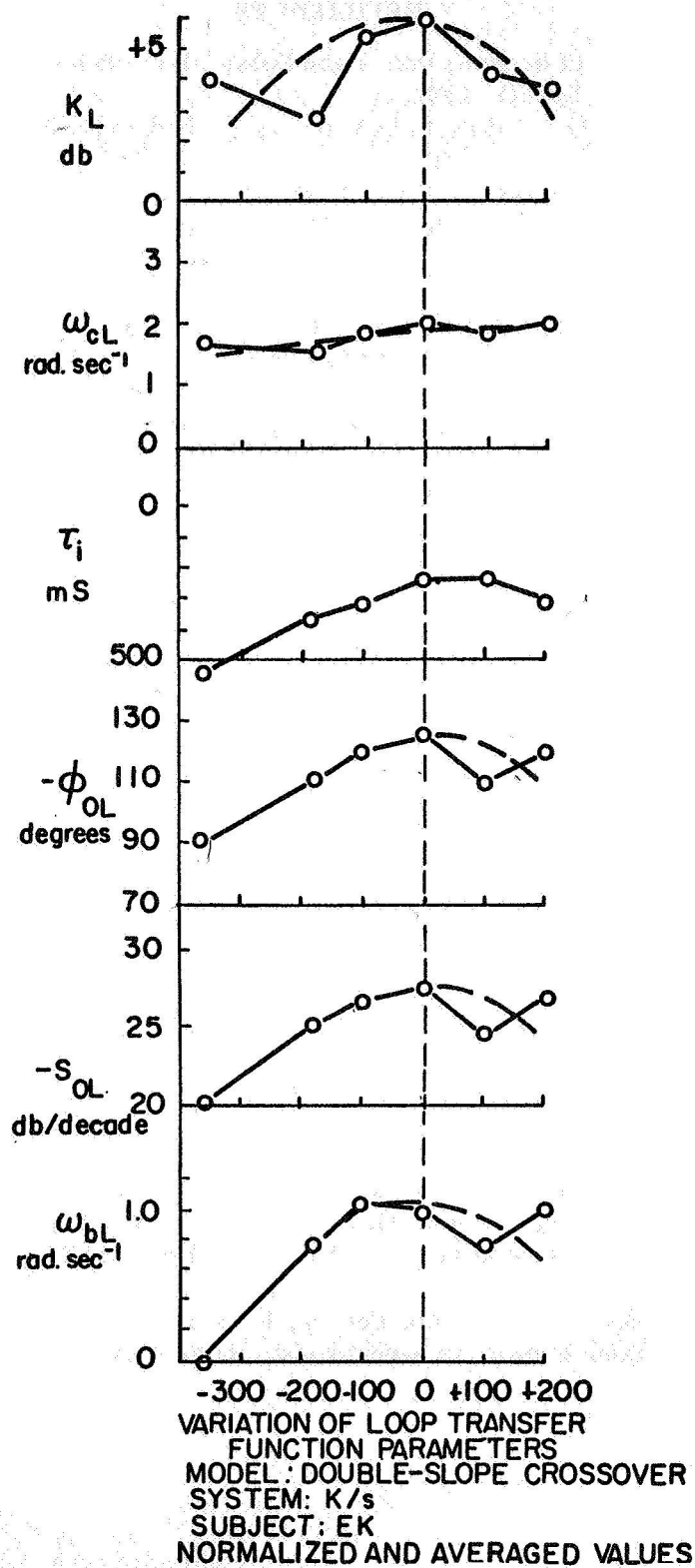


Figure 12.—Double-slope crossover model parameter variations as a function of relative compensation delay.

REFERENCES

1. Nielsen, G.: Control of Systems with Time Delay. Servolaboratoriet (Denmark's Tekniske Højskole, Lynby), 1967.
2. Mufti, I. H.: The Stability of Systems with Lag. Canadian Natl. Res. Council Rept. no. MK-16, Mar 1965.
3. Ogurztoreli, N. N.: Time-Lag Control Systems. Academic Press (New York), 1960.
4. Choksky, N. H.: Time Lag Systems—A Bibliography. IRE Trans. Auto Control, AC-5, Jan. 1960, pp. 66-70.
5. Buckley, P. S.: Automatic Control of Processes with Dead Time. Automatic and Remote Control. Vol. 1., (Proc. IFAC Congress, Moscow, 1960), pp. 33-40.
6. Bhatt, S. J.; and Hsu, C. S.: Stability Criteria for Second-Order Dynamical Systems with Time Lag. Trans. ASME, - J. Appl. Mech., 1965, pp. 1-6.
7. Mori, G.: Discrete Compensator Controls Dead Time Process. Control Eng., Jan. 1962, pp. 57 et seq.
8. Smith, O. J. M.: A Controller to Overcome Deadtime. ISA J., vol. 26, no. 10, Oct. 1965.
9. Lupfer, D. E.; and Oglesky, M. W.: Applying Dead-time Compensation for Linear Predictor Process Control. ISA J., 8, no. 11, Nov. 1961.
10. Gilor, W.: Optimized Feedback Control of Dead Time Plants by Complementary Feedback, IEEE Trans. Auto. Control, AC-9, May 1964.
11. McRuer, D. T.; and Jex, H. R.: A Review of Quasi-Linear Pilot Models, IEEE Trans. Human Factors in Electronics, HFE-8, Sept. 1967, pp. 231-249.
12. McRuer, D. T.; and Krendel, E. S.: Human Pilot Dynamics in Compensatory Systems. AF Systems Command Tech. Rept. AFFDL-65-15, May 1965.
13. Adams, J. L.: An Investigation of the Effects of Time Lag due to Long Transmission Distances upon Remote Control. Dept. of Mech. Eng., Stanford University, NASA Contract NSG 111-61 Progress Rept. no. 3, Nov. 1961.
14. Muckler, F. A. and Obermeyer, R. W.: Control System Lags and Man-Machine Performance. NASA CR-83, July 1964.
15. Hampton, R.; Korn, G. A.; and Mitchell, B.: Hybrid Analog Digital Random-Noise Generation. IEEE Trans. Electronic Computers, EC-12, 412-413, Aug. 1963.
16. McDonnell, J. D.; and Jex, H. R.: A "Critical" Tracking Task for Man-Machine Research Related to the Effective Operator's Delay Time - II: Experimental Effects of System Input Spectra, Control Stick Stiffness and Controlled Element Order. NASA CR-674, Jan. 1967.
17. Wasicko, R. J.; McRuer, D. T.; and Magdaleno, R. E.: Human Pilot Dynamic Response in Single-Loop Systems with Compensatory and Pursuit Displays. Air Force Flight Dynamics Laboratory (AFSC) Tech. Rept. AFFDL-TR-66-137, Dec. 1966.
18. Pew, R. W.; Duffendack, J. C.; and Fensch, L. K.: Sine-Wave Tracking Revisited. IEEE Trans. Human Factors in Electronics. HFE-8, June 1967, pp. 130-134.

15. Psychological and Physiological Skill Development—A Control Engineering Model*

Ezra S. Krendel
University of Pennsylvania
Duane T. McRuer
Systems Technology, Inc.

The application of servomechanism models to the description of human control behavior in either manual or physiological control has generally proceeded by creating circumstances under which behavior would be time stationary in an engineering sense over the duration of the measurements. Remarkably effective stationary steady-state models were achieved and applied with considerable success to the design of complex man-machine control systems. Little attention was paid to defining the behavior whereby the human operator, acting in concert with a machine, achieves higher levels of performance as training progresses. An early attack on this question was made by Krendel and McRuer in 1960. The problem is that conventional performance measures for human manual control or physiological control, interpreted in an invariant closed-loop configuration, are either relatively insensitive to human learning behavior or are inappropriate and misleading. A restructuring of the control configurations to enable the extracting of coherence and regularities from the input structures, either physically or conceptually, is the core of the model presented. Experimental data from many sources are presented to substantiate the model, and implications for system design and training are explored.

INTRODUCTION

In 1960, Krendel and McRuer (ref. 1) presented a hypothesis for motor skills development which was compatible with the current state of knowledge in manual control and human skills training. This hypothesis held that improved human motor skill performance occurred as the operator became successively more proficient in sensing the coherence in the input signal that he was controlling. It was further maintained that the human restructured his effective control configuration to exploit the coherence that he had detected. Three control configurations were specified: a compensatory or error correcting only mode, a pursuit or target following mode, and finally a precognitive or synchronous generator

*The work reported here was sponsored by the Ames Research Center of the National Aeronautics and Space Administration under Contract No. NAS2-2824.

mode. In the final mode preprogrammed patterns of response are generated and timed appropriately. This hypothesis was supported by experimental evidence. In the present paper, the hypothesized configurations have been elaborated, additional empirical evidence gathered by various researchers is presented, and the dependencies of pattern generation on human precision time estimation is illustrated.

GLOSSARY

$c(t)$	operator output time function, limb position
$C(j\omega)$ or C	Fourier transform of operator output
$e(t)$	error time function
$i(t)$	forcing function time function
$I(j\omega)$ or I	Fourier transform of forcing function
K	open-loop gain
K_c	controlled element gain
$m(t)$	system output time function
$M(j\omega)$ or M	Fourier transform of system output
$n_c(t)$	operator remnant time function
$O()$	order of ()
$Y_c(j\omega)$ or Y_c	controlled element (machine and display) transfer function
$Y_p(j\omega)$ or Y_p	pilot describing function
ζ	damping ratio
σ	standard deviation
τ_e	effective time delay
Φ_{im}	cross power spectral density between i and m
Φ_{ie}	cross power spectral density between i and e
$\Phi_{ii}(\omega)$	forcing function power spectral density
ω	angular frequency, rad/sec
ω_c	system crossover frequency, i. e., frequency at which $ Y_p Y_c = 1$
ω_n	undamped natural frequency
\doteq	approximately equal to
\hat{x}	estimate of x

AN ENGINEERING MODEL FOR SKILL DEVELOPMENT

INTRODUCTION.—Skill development, or any learning model, implies a change in performance characteristics as time passes and experiences aggregate. For such an adaptation to occur, the system must be able to sense the relevant aspects of its performance,

compare these with some ideal or selected criteria and, within the limits imposed by the criteria and physical realizability, modify its characteristics. It has been popular to suppose that attainment of satisfactory accuracy levels, while minimizing operator effort, are among the criteria usually involved. Indeed, it has been demonstrated that the human operator can adjust his describing function parameters to achieve minimum or low values of relative rms tracking error (ref. 2), but this performance measure may not always be the sole or even the dominant criterion to which the operator's efforts are directed.

In such an event, improvements in skill could not be measured properly by relative rms error alone and would have to be implied from describing function measurements, physiological measurements, and perhaps operator opinions obtained under carefully defined conditions, the foregoing measures taken separately or together. These measurements directly indicate changes in the operator's behavior, whereas rms tracking error may not. This point will be made clearer later in the context of our skill development models.

As the starting point in model development, let us assume that the human operator is attempting to maintain a small, visually sensed error $e(t)$ or to minimize some function of $e(t)$ while tracking random or random-appearing inputs in a conventional compensatory closed-loop configuration. In figure 1, this would be represented by the system elements connected with heavy lines. A considerable body of findings (refs. 2 to 4) has shown how the parameters of the operator describing function Y_{pe} , which acts on the error signal only, are modified for different controlled elements Y_c and then adjusted during the course of learning to improve performance. In the simplest terms, the operator describing function adapted for a particular controlled element is such that the open-loop system describing function in the neighborhood of system crossover frequency ω_c approximates:

$$Y_{pe}Y_c \doteq \frac{\omega_c}{j\omega} e^{-j\omega\tau_e} \quad \omega = 0(\omega_c) \quad (1)$$

after adaptation of an appropriate Y_{pe} form, the operator can improve system rms error performance by decreasing effective time delay τ_e and extending crossover frequency ω_c thereby increasing the system bandwidth and closed-loop damping. The amount of adjustment possible is limited by operator training and physical limitations, and system stability considerations. Then, once the adapted form for Y_{pe} is optimized by parameter adjustment, the only route that allows further improvement in performance is to achieve different effective adapted forms by either selective loop closures or skill data processing.

PLAUSIBLE LOOP CLOSURES AND DATA PROCESSING. — Because both selective loop closures and data processing will be achieved by changes internal to the human operator, the nature of such changes is conjectural. One can, however, examine performance measures, open- and closed-loop describing functions, and other selected measures for control configurations that are in physical fact elaborations on the simple single-input, single-output unity feedback system. By comparing such measures with similar measures from a subject constrained to operate in what is physically the rudimentary, heavy line system of figure 1, we can imply functional correspondences between the physical elaborations and the internal developments associated with increasing skill.

In figure 1, the elaborations, beyond adaptive and optimizing modifications in Y_{pe} , proceed by the succession of positions of switches α and β . Switch α is in the down position, except when β is on position 2, in which case α is also on terminal 2. When only a compensatory display is present, the informational bases for the elaborations are displayed error e and operator's output c as measured by joint and other proprioceptive

receptors. As a first step, this information is sufficient for the operator to have an internal measure of his own behavior; i. e., Y_{pe} , as a separate entity.

After some experience in controlling a given set of controlled element dynamics, there must be some measure of Y_c within the operator. For example, in terms of the crossover model, at a particular level of Y_{pe} adaptation, an estimate, \hat{Y}_c , for Y_c would be:

$$\hat{Y}_c = \frac{\omega_c e^{-j\omega_c \tau_e}}{j\omega Y_{pe}}, \quad \omega = 0(\omega_c) \quad (2)$$

With this measure, and the output, $c(t)$, it is possible that the operation illustrated in the short-term memory (STM) block takes place, thus enabling a separation from the error signal of either an estimate of the input $\hat{i}(t)$, or of the input of the controlled element, $\hat{m}(t)$. Such an additional effective input can then be operated on by additional blocks Y_{pi} or Y_{pm} . The result is an addition to the output of the heavy-lined system given by

$$\Delta C = Y_{pi} \hat{I} + Y_{pm} \hat{M} \quad (3)$$

This is the pursuit level of skill development. A further possibility is that rather than operating on $\hat{i}(t)$ or $\hat{m}(t)$, the input estimate is scanned for coherence—conceptually perhaps by comparing $\hat{i}(t)$ with a "pattern file" in the long-term memory (LTM). Once a pattern is established, a free-running pattern generator could be triggered setting off an open-loop output of patterned responses via terminal 2. This is the precognitive level of skill development.

THE PURSUIT DISPLAY.—In an actual pursuit display (fig. 2), system output m and forcing function i are presented concurrently on a display. Error e may be extracted by the observer since $e=i-m$. Although only two of these functions are independent, three operations, symbolized by Y_{pi} , Y_{pe} , and Y_{pm} are conceivable. However, no more than two are required for an analytic description of the system. For convenience, two are selected and an analysis is carried out as if they remain fixed choices. This may not be the actual mode of operation. For example, the operator may cycle his attention among the three alternatives, perhaps emphasizing two, and relying on his STM to provide estimates of the other quantities to which he is not attending. The structure labeled A and STM represents this selective attention and STM function. In figure 2, the operator's control output c contains both the results of linear operations and an additive remnant term n_c uncorrelated with i . (For simplicity, the remnant was neglected in connection with fig. 1.)

The closed-loop describing function for any utilization by the operator of the possible feedbacks and feedforward in figure 2 is M/I . It would be expected that M/I would provide a more dynamic picture of skill progression than average performance (e. g., $\overline{e^2}$) alone. Since a closed-loop description tends to suppress internal details, Wasicko et al. (ref. 5) have demonstrated how an equivalent open-loop describing function Y_β can be defined after carrying out block-diagram algebra on figure 2.

$$Y_\beta = \frac{\Phi_{im}}{\Phi_{ie}} = \frac{Y_c(Y_{pi} + Y_{pe})}{1 - Y_c(Y_{pm} + Y_{pi})} \quad (4)$$

$$M/I = \frac{Y_\beta}{1 + Y_\beta} \quad (5)$$

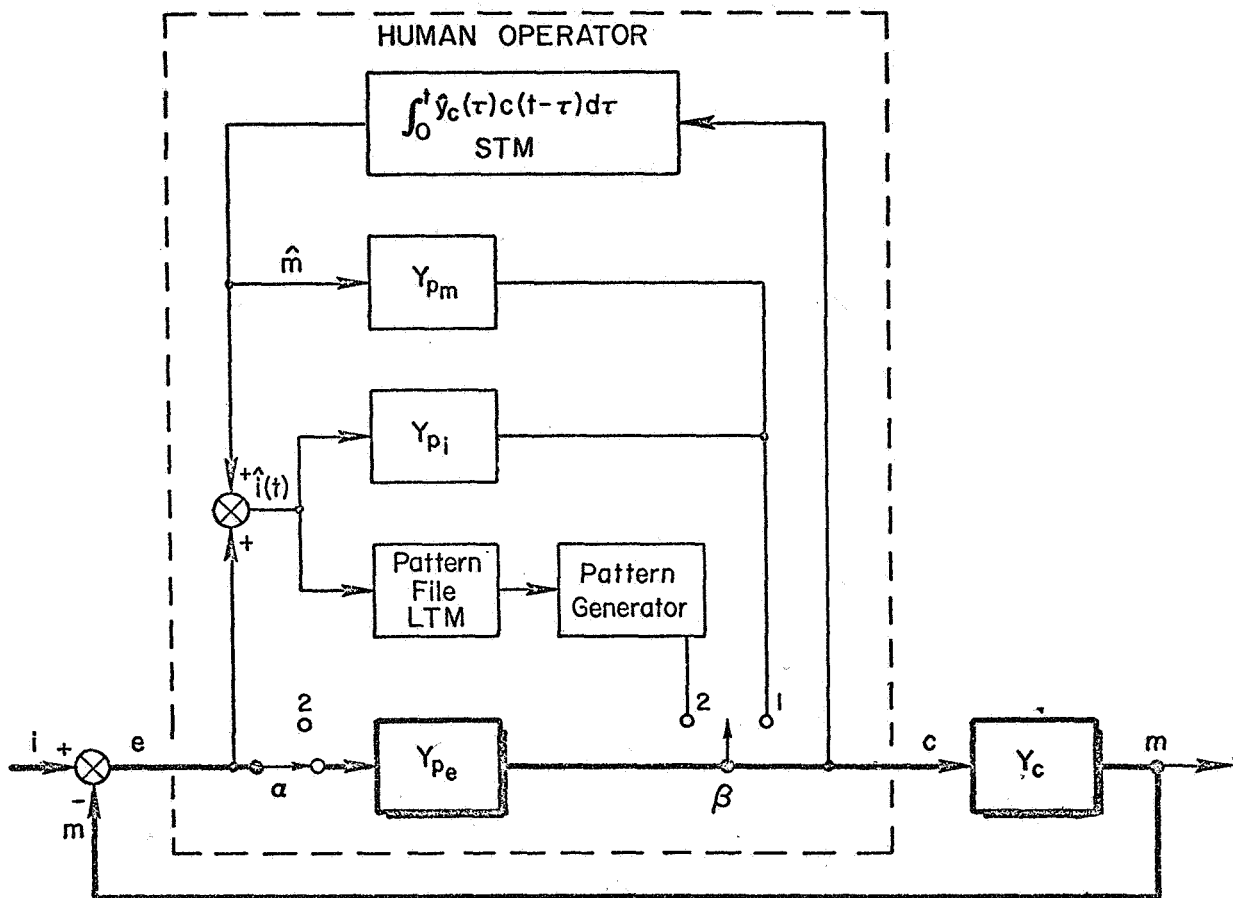


Figure 1. — Functional description of skill development—SOP.

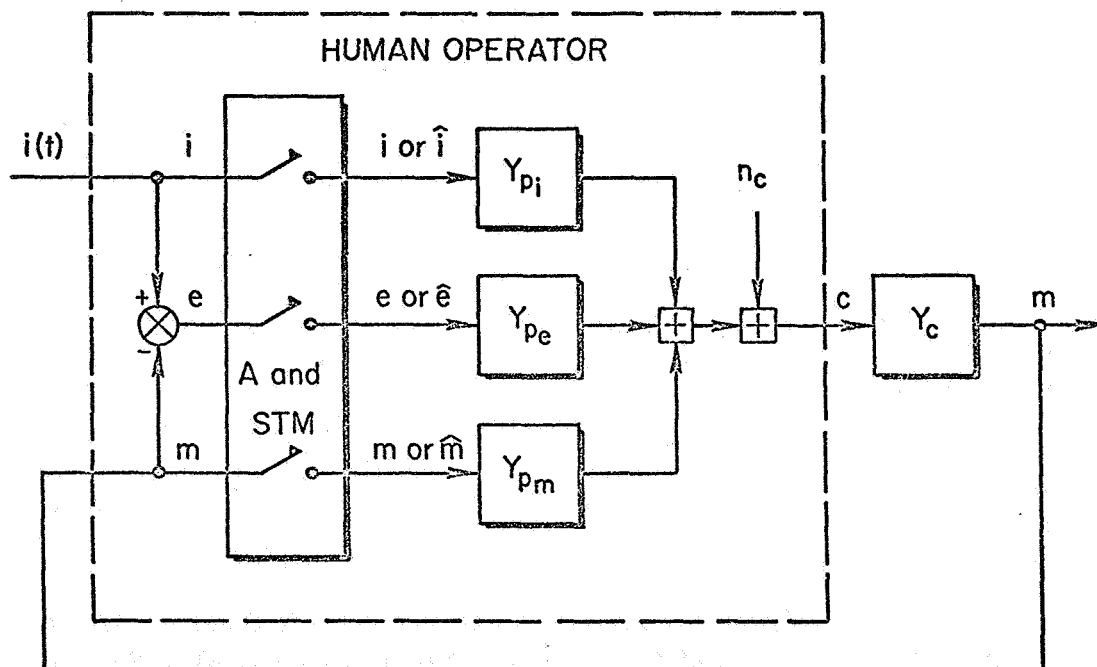


Figure 2. — The pursuit situation.

If the situation is compensatory; i. e., $Y_{p_i} = Y_{p_m} = 0$, the equivalent open-loop describing function becomes simply Y_{p_e} .

The pursuit display gives the operator a greater repertory from which to adjust Y_{p_e} , so as to improve performance. The mean squared error is given by two terms: $\overline{e^2}$ due to operations linearly correlated with the forcing function; and e_n^2 due to the remnant.¹

$$\overline{e^2} = \overline{e^2}_i + \overline{e^2}_n \quad (6)$$

where
$$\overline{e^2} = \frac{1}{2\pi} \int_0^\infty \left| \frac{1}{1 + Y_{p_e}} \right|^2 \Phi_{ii}(\omega) d\omega, \quad (7)$$

From equations (4) and (7), the implications for the adjustment of the describing functions that comprise Y_{p_e} can be drawn. The operator must maintain a Y_{p_e} block if only to trim out drifting errors in figure 2. He may emphasize Y_{p_i} or Y_{p_m} to the practical exclusion of the other, or may cycle between them in some weighted fashion. The operator can improve some measure of average performance by increasing the effective open-loop amplitude ratio Y_{p_e} thereby increasing the effective open-loop crossover frequency. For instance, Y_{p_e} can be increased if:

$$Y_{p_i} \approx 0 \text{ and } Y_{p_m} \approx 1/Y_c \quad (a)$$

or
$$Y_{p_m} \approx 0 \text{ and } Y_{p_i} \approx 1/Y_c \quad (b)$$

Wasicko et al. (ref. 5) solved equation (4) for $Y_{p_i} Y_c$ using the assumptions that $Y_{p_m} = 0$, and that Y_{p_e} was the same as that which would have been generated in an actual compensatory display with the same $i(t)$. For controlled elements, where pursuit behavior was demonstrated; i. e., where $Y_{p_e} \neq Y_{p_e}$, the results are most consistent with (b) above in the region of crossover.

ULTIMATE SKILL—PRECOGNITIVE.—The pattern file, pattern generation precognitive configuration in figure 1 represents the ultimate in skill development because in it the operator can construct an estimate of the future input. With this information, he can compensate for lags inherent in the human neuromuscular system and the controlled element, thereby making the output duplicate the forcing function with no lag or even some apparent lead. Precognitive organization, because of its anticipatory nonlinear and open-loop nature (Y_{p_e} is only occasionally closed for trimming) is difficult to describe with conventional control engineering techniques. Figure 3 is a schematic for the precognitive mode of skill. In a formal manner, verbal descriptions of precognitive behavior, "deft, properly timed, appropriate loop closures, etc.," can be converted to computer-oriented flow charts and decision trees. This procedure can make the decision rules in figure 3 more explicit.

The essence of the precognitive mode is the ability to construct the future of the input signal. Examples occur in almost all of our skilled actions—in driving down a road, in docking a boat, and in signing our names, as van der Gon and Thuring (ref. 6) demonstrated. The high degree of reproducibility in a signature or holograph and the obvious lack of continuous visual feedback make it clear that writing is a programed, open-loop skill.

The precognitive mode has been studied most effectively, not in terms of preview or of everyday skills, but by means of periodic inputs that permit the description of prediction in an unambiguous manner. McRuer and Jex (ref. 7), using the available data on periodic

inputs, derived a quantitative description of the evolution of precognitive behavior in an actual compensatory display.

The process can be understood by reference to figure 1, although a remnant or noise source must be added to the operator's output $c(t)$. McRuer and Jex (ref. 7) present their case as follows.

(1) For periodic inputs up to about 10 rad/sec (2 hertz), the operator creates a pursuit display feedforward loop in which $Y_{pi} Y_c = 1.0$, which serves to eliminate all of the time-averaged lags from his tracking. However, remnant will still excite residual errors, so a compensatory loop is also maintained, subject to similar laws of adaptation as in the random input case for very low frequency inputs. The result will be one peak in the perceived error spectrum near the input frequency and, possibly, a secondary peak near the crossover frequency.

(2) As familiarity with a particular waveform increases, Y_{pi} takes on the form of a synchronous waveform generator that is phase locked to the zero crossings of the input. Thus, the operator's behavior becomes more and more open loop or preprogramed, and the errors at the input frequency will decrease.

(3) For inputs between about 10 and 30 rad/sec (2 to 5 hertz), the operator can no longer close the compensatory loop effectively, so crossover regression (or even complete opening of the compensatory loop) occurs. Even though the synchronous generator produces a reasonable waveform, it is no longer phase-locked to the input frequency, so shifts in phase or frequency drifting can occur.

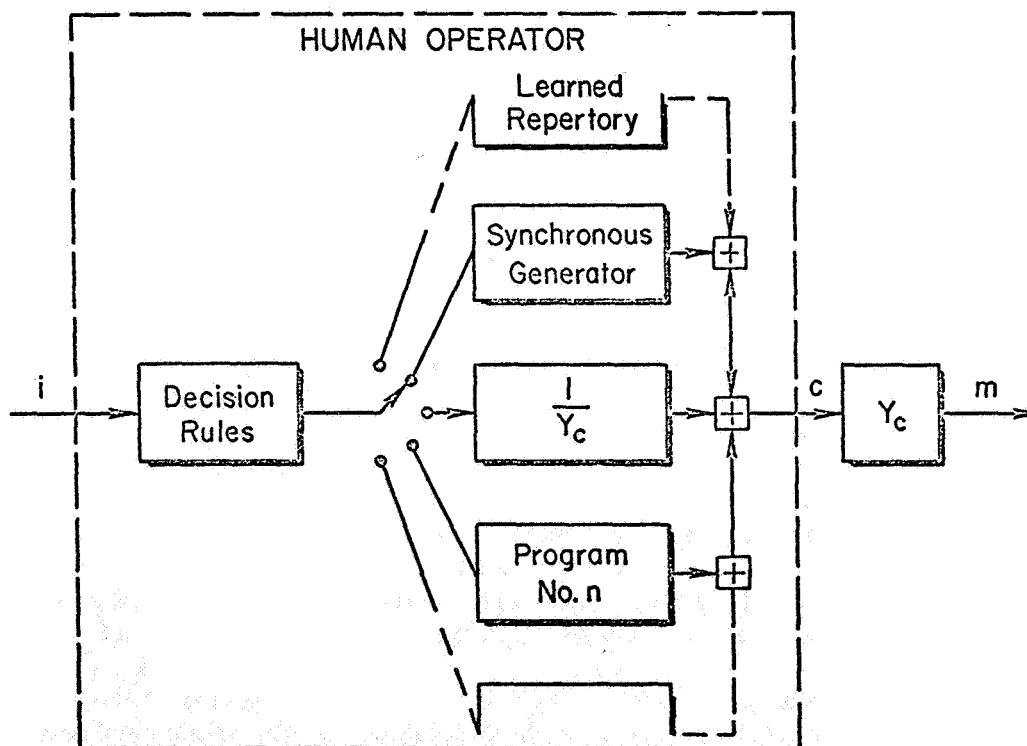


Figure 3.—Precognitive organization—ultimate skill—open loop for varying durations.

(4) The maximum periodic frequency is limited by the basic neuromuscular dynamics of the limb plus control stick, which also limits the sharpness in reproducing periodic square or triangular waveforms.

The rapidity with which most operators develop a synchronous output, at least when $Y_c = K_c$, makes it likely that phase b in the previous progression takes place within two or three cycles. This means that it is of considerable interest to discover the conditions, limits, and accuracies with which periodic inputs can be reproduced, maintained, and outputs phase-locked with these inputs.

It is well to reiterate that the schematics in figure 1 and figure 3 refer to either actual or effective display configurations. Thus, it is possible in a physical pursuit configuration for the operator to perform as if the display were, in fact, compensatory; whereas after much practice with a physically compensatory display on an input which is predictable, the operator can achieve an effective pursuit or precognitive display (refs. 8 and 9). It is the operator's level of skill, not the physical display-control configuration that determines the characteristic mode of behavior in the successive organization of perception (SOP) sequence.

PSYCHOLOGICAL AND PHYSIOLOGICAL CONSIDERATIONS

In this section, we describe some signal processing characteristics of the human which define the limits of his ability to acquire and to perfect motor skills. In the context of the engineering model, these are characteristics pertaining to pursuit and precognitive operation in limiting conditions. Because our point of view is both unifying and novel, the data we present are from diverse sources and are not complete.

LEARNING THEORIES.—Much of the traditional psychological literature on motor skills is either specific to a particular task or group of tasks or deals with observed quantities that are not particularly useful in a control engineering connection. If, however, we consider psychological learning theories more generally; i. e., as they were developed for more extended applications than to motor skills alone, we will note an emphasis on concepts that are highly relevant to a control engineering model. One can describe learning theories along a continuum anchored on one extreme in an emphasis on stimulus organization (the Gestalt theories) proceeding through a range of association and classical conditioning theories that emphasize stimulus-response time intervals to another anchor in Skinnerian (ref. 11) operant conditioning which emphasizes the organization of the responses.

The role of the stimulus-organization in learning has been emphasized by the Gestalt psychologists (e. g., Koffka, ref. 10) and expressed in terms of concepts such as closure, which approach the esthetic. Learning proceeds by insights attributable to a perceptual reorganization of the stimulus field which makes the problem's solution clear. In motor skills learning, we must define perception more broadly, so that in addition to conventional verbal responses, motor responses, (for example, manual control) can manifest or modify a perception or pattern in time or space. This emphasis on the structure of the input is shared by coaches and physical trainers. Trainers are interested in gross motor ability rather than in the fine muscular activity with which most human control behavior studies are concerned; however, the pragmatic attitude of the trainer forces on him insights that often elude others. For example, Harrison (ref. 12), in a detailed review of the neuromuscular bases for motor learning, comes to the conclusion that the prime emphasis in training should be put on organizing sensory inputs. This was expressed as follows: ". . . the role of the teacher is to give the chief sensory cue to the learner to use at the right time and let nature take its course . . ."

In terms of our model, the teacher serves as a device that first separates the input signal from input noise, and then emphasizes the predictability of the signal so that the trainee can select the proper output from his repertory of preprogramed responses.

To describe motor skills development, we must examine response as well as stimulus organization. Skinner, in his emphasis on emitted responses, has attacked this end of the problem. Skinner's emphasis on the response arises because, although one may not be able to isolate the specific stimulus that triggers a particular response, one can describe the regularities intrinsic in the patterns that the responses themselves define. It is as if we were describing the organism in terms of its ability to act as a free-running generator of responses, and only secondarily, if at all, addressing the phase synchronization of these responses with a stimulus or stimulus pattern. In the context of our model on figure 1, he has treated the rules that define how response patterns become incorporated into the pattern generator, and how they can be modified.

Insofar as motor skill is an example of the effective functioning of a control system, a synthesis of the foregoing polarized points of view is possible within a control engineering model.

RHYTHMIC TAPPING.—The data (refs. 13 to 16) which we examine will use highly organized inputs and responses (periodic functions in both cases) to manifest the limits of predictive human motor behavior.

Tapping in response to periodic inputs is a human output for which only the phase characteristic is definable. In this sense, it is a simpler task than conventional tracking problems. By examining the tapping literature, we can get information on two characteristics of human prediction: the extent to which prediction is, in fact, achieved; and the precision with which a periodic input is reproduced. The first is an equivalent of phase lag or advance. The second is a measure of the bandwidth of the output frequency generated in response to a single frequency input. The tapping equivalent of the effective and actual compensatory display is the conventional reaction time experiment in which a motivated subject attempts to minimize the delay between his response and the signal. A fully trained and altered subject produces a reaction time delay and associated variance which is a measure of neuromuscular limitations. For taps in response to a visual input, a 180 ms delay and a standard deviation of 15 to 20 ms are typical values. A subject tapping in an attempt to synchronize with a periodic signal represents an equivalent of a pursuit mode since the input rapidly becomes known to the subject. In general, synchrony is achieved within three or four periods. Characteristically, anticipation occurs as illustrated in figure 4.

At least as important a measurement, from the viewpoint of a subject approaching the precognitive mode of behavior, is the quality with which a periodic input is reproduced. The standard deviation σ of the stimulus-response interval for a periodic stimulus provides this measure. As shown on figure 5, the product of σ and input frequency $f=1/T$ is a constant of the order of 0.04 to 0.06 up to a minimum period T of 200 to 400 ms. At this point, frequency drifting begins and the product rapidly increases to a limiting value of $1/\sqrt{12}$. This limit represents a condition where the subject's response is selected from a rectangular distribution over the interval between stimuli T . The time interval 200 to 400 ms may represent a lower limit of a memory function that is able to maintain a constant σ/T up to this time interval.

PATTERN GENERATION.—Tracking data for square waves and sine waves are consistent with figure 5 (e.g., ref. 17). Complex combinations of sine waves have also been

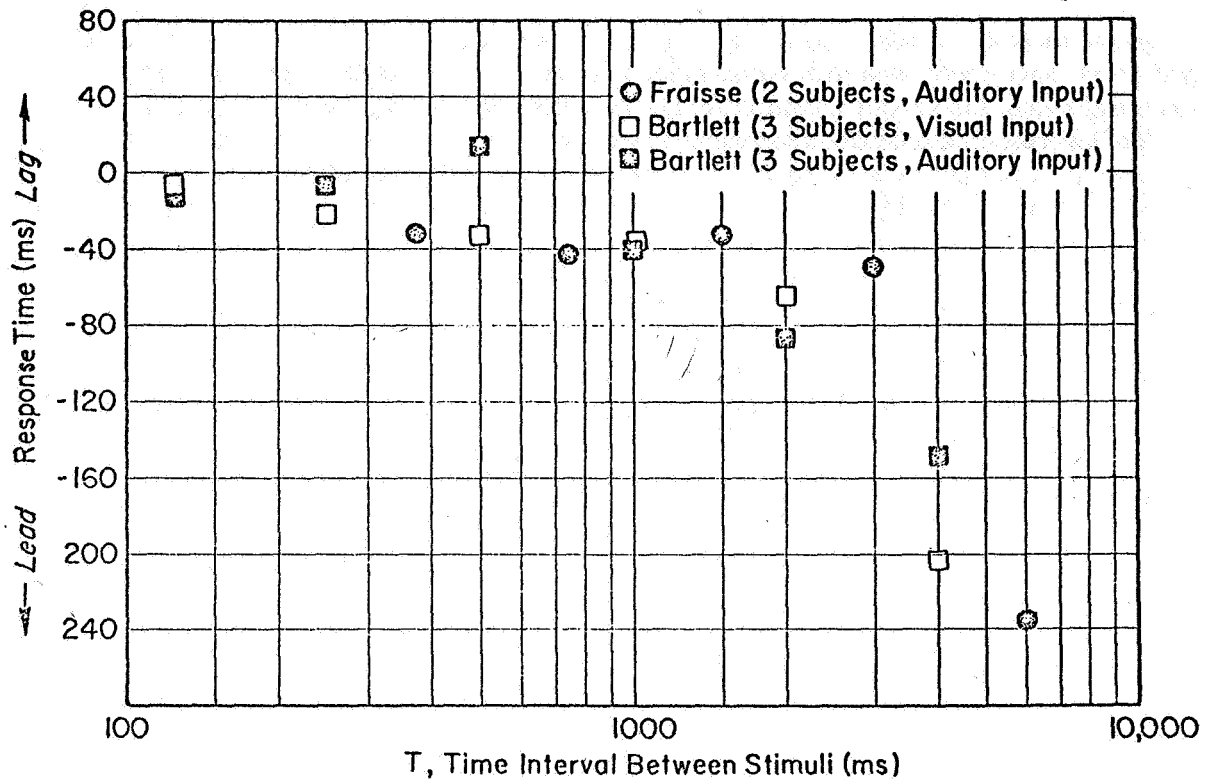


Figure 4.—Synchronization as affected by stimuli interval.

used to demonstrate the operator's ability to generate patterns and synchronize them. Vossius (ref. 9) created a periodic function of frequency 0.75 hertz, which consisted of a fundamental and its first two harmonics. This function was the input to a subject in an actual pursuit configuration. After a training experience of 20 to 30 periods, the input was removed during the course of pursuit tracking and the subject continued to recreate the input from memory. From figure 6, two points are clear: The open-loop generated pattern is not in phase with the input, and the open-loop pattern has a period that is about 10 percent longer than that of the closed-loop output.

Hess (ref. 18) demonstrated a similar finding in a study of human control of periodic, time-varying, plant dynamics. Using an actual compensatory display, a single frequency input of 1 rad/sec, and plant dynamics $Y_c(4) = 9.66 (1 + 0.5 \cos 4t)$, he trained an operator for about 40 test runs and then blanked the error display for 8 to 9 seconds. Figure 7 indicates this blanking by double-headed, horizontal arrows. The stick movement is clearly unchanged in pattern, as in figure 6, but there appear to be phase and frequency changes. Clearly, the operator is not attending to the error in a closed-loop fashion. He is operating in a pattern generating or precognitive mode.

Pattern generation may not necessarily result in improved performance; however, if the major system consideration is that of stabilizing a divergent controlled element, the

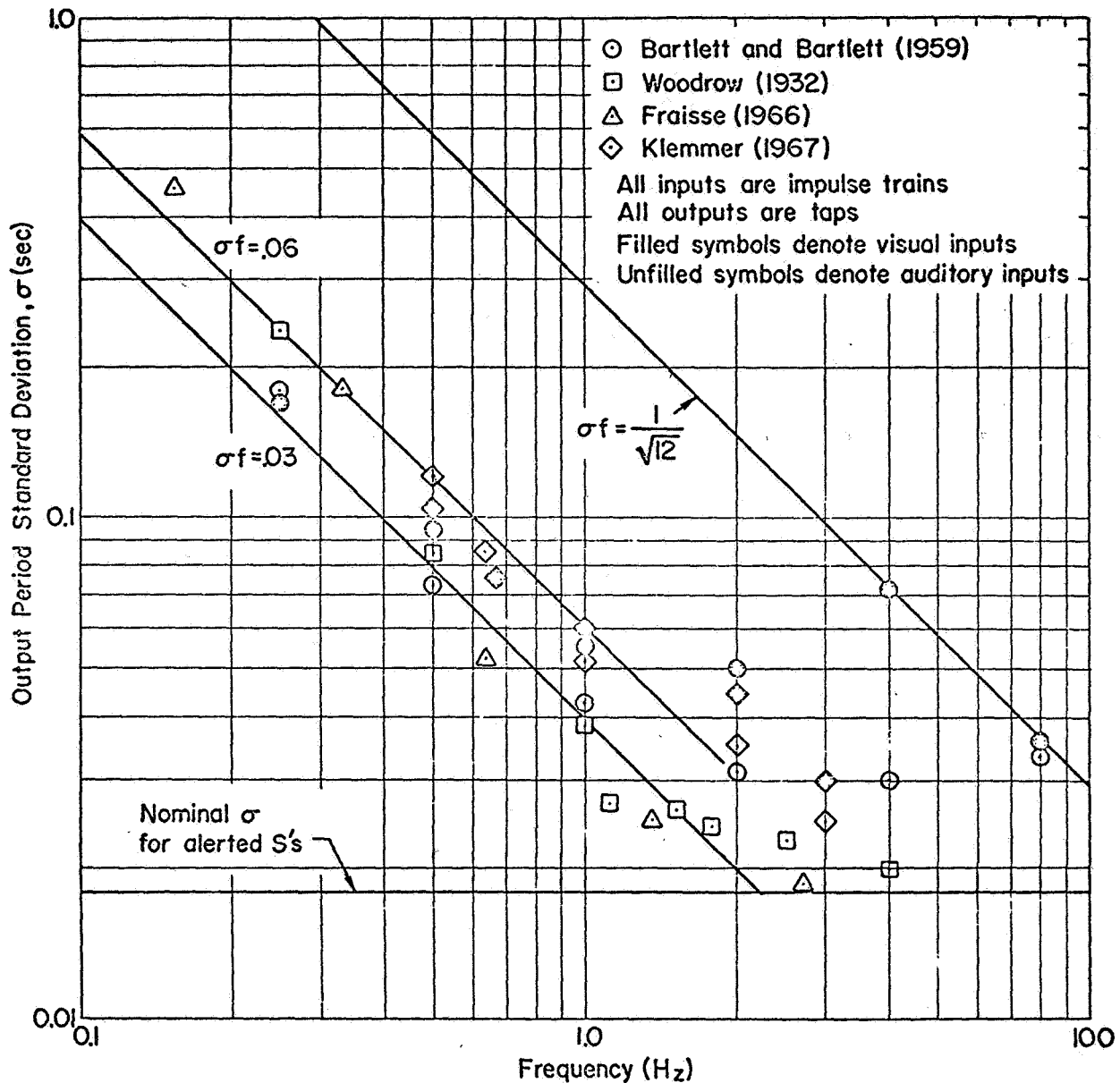


Figure 5. —Standard deviation of general period versus frequency of input function.

ability to predict may be critical. Smith (ref. 19) examined a human operator stabilizing a second order of oscillatory divergence. An actual compensatory configuration was used and the controlled element was:

$$Y_c(s) = \frac{K_c}{s^2 + 2\zeta_n \omega_n s + \omega_n^2}$$

where ζ_n is negative.

The experimental procedure consisted of presenting two subjects with an input consisting of four nonharmonic sine waves, 0.28, 0.74, 1.21, and 1.80 rad/sec, selecting a

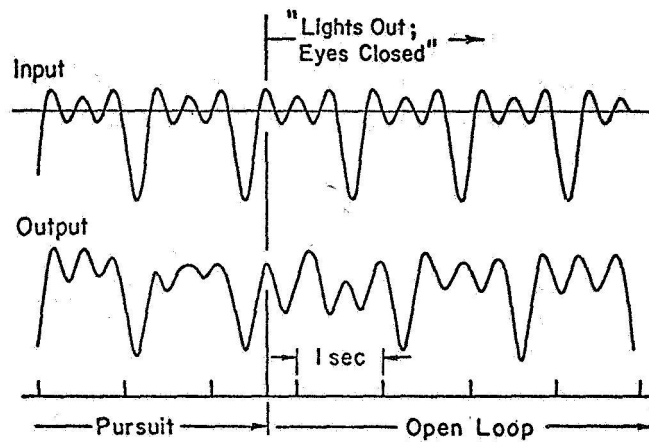


Figure 6.—Demonstration of synchronous pattern generation.

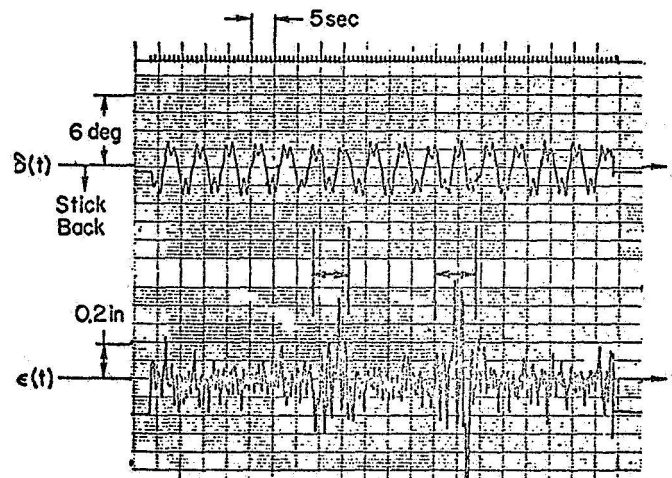


Figure 7.—Synchronous pattern generation stick output and error signal for oscilloscope blanking experiment.

value of ζ_n equal to -0.35 or -0.7 and increasing ω_n by discrete amounts on successive runs from 2.5 to 7.2 rad/sec until a maximum controllable value was achieved and maintained for a 2-minute run. It required 392 attempted runs of which 131 were successful, i. e., tracking was maintained for 2 minutes to establish controllability limits for $\zeta_n = -0.35$ over the frequency range measured.

After this extensive training, Smith's subjects stabilized a divergent second-order plant, in an actual compensatory display, with a random appearing input where stability would not have been predicted from our knowledge of the possible characteristics of Y_{pe} . The effective delay time, T , for example, would have had to be of the order of the synaptic delays, or 20 ms. The trackers must have established a more effective control structure than the compensatory mode. Precognitive organization, by enabling the tracker to compensate for his inherent lags, made stable control realizable.

CONCLUSION

Starting with the most primitive configuration in which the human operator is a single input, single output element in single feedback control system, the familiar compensatory mode, a progression in system configurations explains how improvements in skill can arise in a servomechanism context. The pursuit mode, either by an actual physical configuration or as an effective reorganization as a result of learning, provides a mechanism whereby the feedforward necessary for compensating for the neuromuscular lags inherent in a human operator can arise. Ultimate skill arises when either actually or effectively the operator can predict his required control maneuvers and has the capability to execute them. The actual configuration can provide preview, but it is in modifying the effective configuration that the greatest challenge lies. In the effective precognitive mode, at a time determined by his prediction of the output signal's form, the operator releases a repertory of patterned responses that he has acquired by extensive experience and training.

When skill development is viewed in the foregoing context, a framework becomes available for including the psychological learning theories that emphasize the order and pattern of the input structure and insights, as well as the approach that emphasizes regularities in the output structure without attempting to isolate the input stimuli specifically. Furthermore, insights into the skill development process became possible which would not arise from more conventional psychological frameworks. For example, the possibility of measuring a short-term memory function for other than verbally mediated events is possible. Similarly, pattern recognition without verbal mediation becomes feasible. Lapses of skill may become understandable once the perceptual and training necessities for each level of skill are specified. Some accidents may be due not to shortcomings in motor control, but to a misapprehension on the subject's part of the control configuration in which he was operating. When we examine the precognitive mode, the evidence from such diverse sources as handwriting, tapping and tracking, and anecdotal studies of athletic skill improvement, all points in one direction. After familiarity with the plant dynamics has occurred and the operator has arrived at a beginning level of skill, further improvements occur as the operator can predict his required control responses. Characteristically, first a patterned output is emitted and then this output is properly phased with the appropriate stimuli.

We are beginning to be able to assign numbers to some of the quantities in this model for skill development. Thus, it appears from a consideration of much diverse data that the frequency spread of human generated single sine waves, as measured by σ of the drifting frequencies, is directly proportional to the interval between the periodic input signals up to a minimum interval T . Thus, $\sigma f = \sigma T = k$; where $k \approx 0.04$ to 0.06 up to an interstimulus interval of 200 to 400 ms, at which point k increases to a limit of ≈ 0.29 . From many sources, there is evidence that the synchronous manual generation of single sine waves can take place after only 3 or 4 input cycles. There is evidence that after hundreds of trial runs, an operator can exploit the regularities of the input and the controlled element to achieve an effective precognitive mode and stabilize systems under conditions previously thought to be impossible.

Much more quantitative data is needed to extend the usefulness of the SOP theory of skill development. With more data we will become able to design for ultimate skill. We will also attain valuable new insights into human information processing behavior.

REFERENCES

1. Krendel, E. S.; and McRuer, D. T.: A Servomechanism Approach to Skill Development. J. Franklin Institute, vol. 269, no. 1, Jan. 1960, pp. 24-42.

2. McRuer, D. T.; Graham, D.; Krendel, E. S.; and Reisener, W., Jr.: Human Pilot Dynamics in Compensatory Systems—Theory, Models and Experiments with Controlled Element and Forcing Function Variations. AFFDL-TR 65-15, Jan. 1965. In abbreviated form, this was published as Manual Control of Single-Loop Systems, J. Franklin Institute, parts I and II, vol. 283, nos. 1 and 2, 1967.
3. McRuer, D. T.; Krendel, E. S.; and Graham, D.: Adaptive and Optimizing Behavior of the Human Operator in Compensatory Tracking. Proc. XVth International Astronautical Congress, Warsaw, 1964.
4. McRuer, D. T.; Graham, E.; Krendel, E. S.; and Reisener, W. C., Jr.: System Performance and Operator Stationarity in Manual Control Systems. Proc. Third IFAC Congress, London, June 1966.
5. Wasicko, R. J.; McRuer, D. T.; and Magdaleno, R. E.: Human Pilot Dynamic Response in Single-Loop Systems with Compensatory and Pursuit Displays. AFFDL-TR 66-137, Dec. 1966.
6. van der Gon, J. J. Denier; and Thuring, J. Ph.: The Guiding of Human Writing Movements. Kybernetik, vol. 2, no. 4, 1965.
7. McRuer, D. T.; and Jex, H. R.: A Review of Quasi-Linear Pilot Models. IEEE Trans., vol. HFE-8, no. 3, Sept. 1967.
8. Sheridan, T. B.: Time-Variable Dynamics of Human Operator Systems. Sc.D. Thesis, MIT, Cambridge, Mass., 1959.
9. Vossius, G.: The Prediction Capabilities of the System of Haphazard Motion. Neuere Ergebnisse der Kybernetik. (Steinbuch, Wagner, Oldenbourg, R., eds.) Munchen, 1964.
10. Koffka, K.: Principles of Gestalt Psychology. Harcourt Brace, 1935.
11. Skinner, B. F.: The Behavior of Organisms. Appleton, Century Crafts, 1938.
12. Harrison, V. F.: A Review of the Neuromuscular Bases of Motor Learning. Res. Quart. Am. Assoc. Health, Phys. Educ. and Rec., vol. 33, no. 1, 1962, pp. 59-67.
13. Woodrow, H.: The Effect of Rate of Sequence Upon the Accuracy of a Synchronization. J. Exptl. Psychol., vol. 15, no. 7, 1932, pp. 357-379.
14. Bartlett, N. R.; and Bartlett, S. C.: Synchronization of a Motor Response with an Anticipated Sensory Event. The Psychol. Rev., vol. 66, no. 4, July 1959, pp. 203-218.
15. Fraisse, P.: L'anticipation des stimulus rythmiques, vitesse d'establissement et precision de la synchronization. Annee Psychol., vol. 66, no. 1, 1966, pp. 15-36.
16. Klemmer, E. T.: Sequences of Responses to Signals Encoded in Time Only. Acta Psychol., vol. 27, 1967, pp. 192-207.
17. Pew, R. W.; Duffendack, J. C.; and Fensch, L. K.: Sine Wave Tracking Revisited. IEEE Trans. Human Factors Electronics, HFE-8, no. 2, 1967, pp. 130-134.
18. Hess, R. A.: The Human Operator as an Element in a Control System with Time Varying Dynamics. Air Force Flight Dynamics Laboratory. FDCC-TM-65-34, June 1965.
19. Smith, R. H.: An Experimental Determination of the Limits of Pilot Controllability for Unstable Oscillatory Second Order Systems, A. E. Thesis, University of Cincinnati, 1961. An abbreviated version is in IEEE Trans. Human Factors in Electronics, HFE-4, On the Limits of Manual Control. Sept. 1963.

16. The Application of Discrete Modeling Elements to the Synthesis and Identification of a Deterministic Model for the Visual Scanning Behavior of Human Operators*

M. J. Merritt

University of Southern California

A deterministic model for the visual scanning behavior of a human operator has been synthesized and partially identified. The applicability and generality of the proportional decision element and the multistate decision element to the modeling of discrete human operator behavior has been demonstrated. The hypothesis is advanced that the operator concentrated on one display, sampling the other only as often as needed. Whether this would have occurred if the operator generated continuous stick motions is impossible to say.

ON THE SCANNING BEHAVIOR OF HUMAN OPERATORS

There are many physical systems in which the operator's control actions are based on visual information obtained from two or more physically separated sources. An aircraft is an excellent example of such a system. If a change in altitude is desired, the pilot must monitor the aircraft's pitch angle, altitude, airspeed, and rate of descent. A model for the behavior of a pilot during an instrument climb or descent must contain two separate systems: a manual control system whose outputs are stick position and throttle position, and a visual scanning system that obtains the information needed by the manual control system. Both systems operate together to fly the aircraft. Since the outputs of the manual control system, stick, and throttle position are deterministically related to the displayed quantities, it seems reasonable to assume that the visual scanning behavior would likewise be a deterministic function of the present and past visual samples of the aircraft's instruments.

Models that describe the visual scanning behavior of human operators have been described in the literature. These models may be separated into three groups:

Group I: Tasks in which the displays presented uncorrelated signals. The operator's eye motions were recorded, but the time histories of the displayed signals were not.

Group II: Tasks in which the display devices presented correlated signals. The eye motions were recorded but the signal time histories were not.

Group III: Tasks in which the display devices presented correlated signals. Both the eye motion time histories and the signal time histories were recorded.

*This research was supported by NASA Grant NGR 05-018-022.

Models for the visual scanning behavior of operators in group I and II tasks have been developed by Senders, Elkind, Stevens, and Smallwood (refs. 1, 2, 3) and Carbonell (ref. 4). A recent study by Levison and Elkind (ref. 5) described a group III task, but no attempt was made to develop a model for the operator's eye motions. In all of the investigations described in the literature, the behavior modeled is the distribution function of the fixation times for each instrument and distribution function of the various interdisplay transitions. The resultant scanning model depends on the long-term properties of the displayed signals: signal mean and moments, or bandwidth. The instantaneous signals viewed by the operator are not utilized.

Models for group III situations have not, to the author's knowledge, been presented in the literature. Thus, the primary objective of this paper is to describe the synthesis and identification of a deterministic model for the scanning behavior of a human operator in a group III task. The model is synthesized by combining two general purpose discrete modeling elements, the proportional decision element (PDE) and the multistate decision element (MSDE). The MSDE, PDE, and their associated identification algorithms are described in reference 7.

THE TRACKING TASK

Terrain avoidance systems and instrument landing systems are primarily pitch axis tasks and often utilize similar display configurations: an attitude display and a cross pointer error display. Although the results described below apply equally well to both, the terrain avoidance task was selected as it did not involve pitch trim or the generation of a glidepath. A block diagram of the task utilized is seen in figure 1, and a view of the manual tracking station is seen in figure 2. The operator sat in a long-range transport pilot's chair and operated a low-inertia, spring restrained sidarm controller. Frequency response problems associated with actual aircraft instruments were avoided by utilizing oscilloscopes with specially prepared edge lighted reticles. The oscilloscope displays are seen in figure 3. The

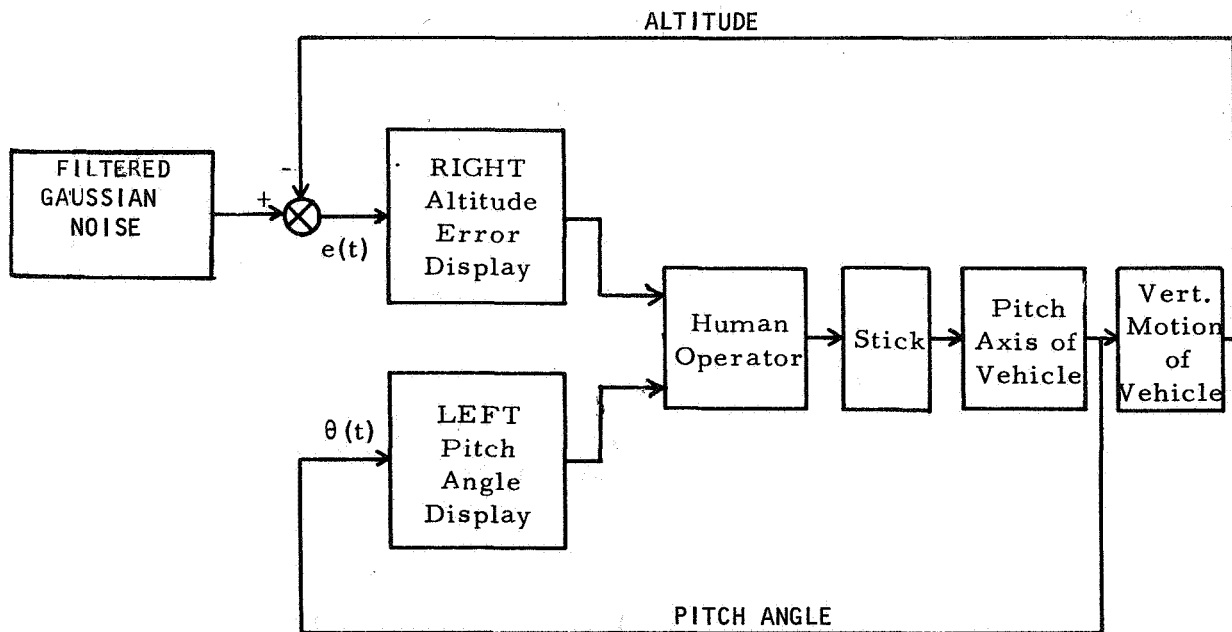
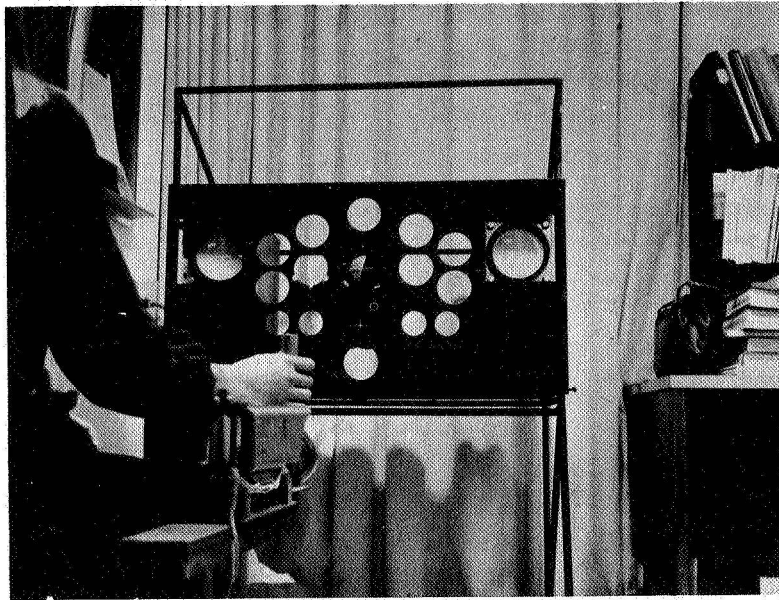
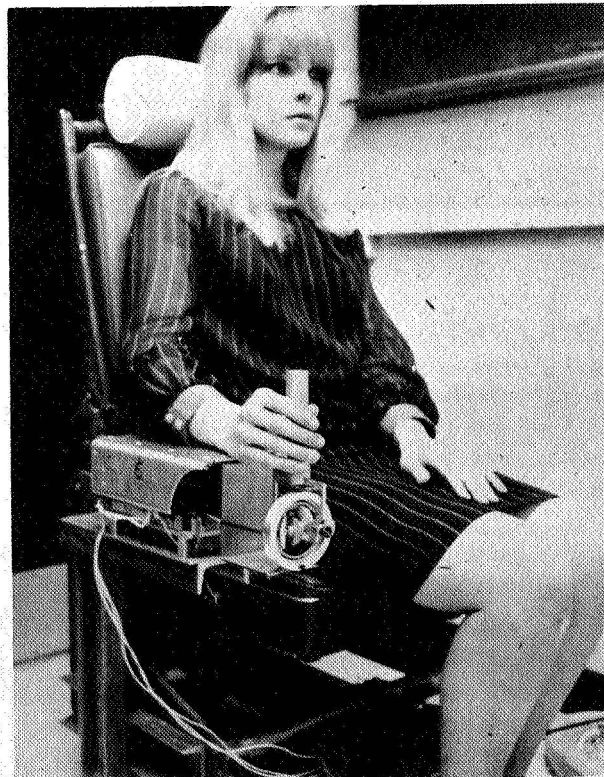


Figure 1. —The tracking task.



(a)

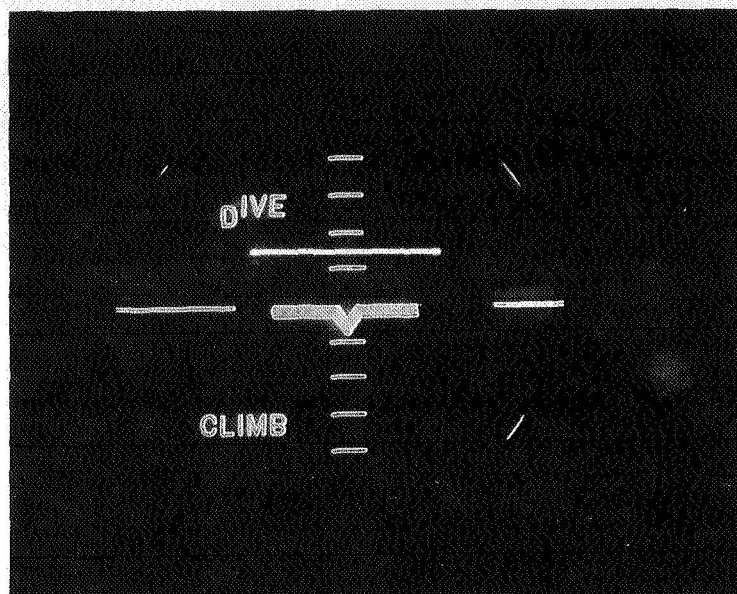


(b)

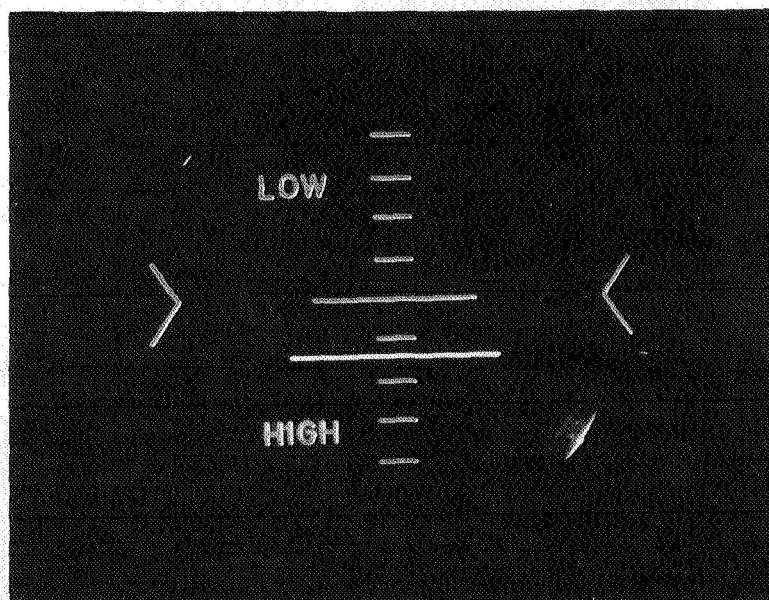
Figure 2. —Tracking station (a) Display configuration,
(b) pilot's seat and sidearm controller.

transfer functions, display signals, and input filter were mechanized on a Beckman 2132 analog computer. Seven signals were recorded on FM magnetic tape and subsequently digitized:

- (1) the operator's eye position
- (2) stick position, $h(t)$
- (3) pitch angle, $\theta(t)$
- (4) pitch rate, $\dot{\theta}(t)$
- (5) altitude error, $e(t)$
- (6) altitude error rate, $\dot{e}(t)$
- (7) input, $r(t)$



(a)



(b)

Figure 3. —Illuminated displays. (a) Left display reticle, (b) right display reticle.

All of the subjects studied were pilots with between 20 and 100 hours of single engine aircraft time. Familiarization with the experimental task was achieved by replacing the gaussian noise source with a constant voltage. This reduced the task to an initial altitude error over a flat terrain and resulted in smooth climbs and descents to zero altitude error. The subjects were instructed to reduce the altitude error to zero as quickly as possible without exceeding plus or minus 40° of pitch angle.

A single subject received approximately 30 hours of training during a 3-week period. One of the last training sessions was sampled by a digital computer at 40 samples per second and stored on a disk file for further processing.

MEASUREMENT OF EYE POSITION USING ELECTRO-OCULAR POTENTIALS

According to Young (ref. 8), a relationship between eye motion and periorbital potential variations was observed as early as 1922. In 1936, Mowrer, Ruch, and Miller (ref. 10) reported that potential differences could be measured between electrodes placed near the eyes that resulted from motions of the eye in the plane formed by the two electrodes and the eye's center of rotation. These potentials were called electro-ocular potentials. The electro-ocular potentials are approximately 20 microvolts per degree of rotation. Although extremely stable, high gain amplifiers are presently available, considerable difficulty is encountered in attempting to measure eye position precisely for extended periods of time. Electrode artifacts, slow biological drifts, etc., all contribute to slow dc offsets. The method is, however, highly suited to the determination of gross eye position for periods of at least 30 minutes. When a small number of displays are widely separated, say 30° or more, no difficulty is encountered in determining which display is being viewed foveally.

The electro-ocular potentials are measured by placing suitable electrodes, preferably Beckman biopotential skin electrodes #350040, laterally as shown in figure 4. The left and right electrodes are connected to the differential inputs of a high gain, high common mode



Figure 4. —Subject showing location of biopotential skin electrodes for eye motion measurements.

rejection, dc amplifier. The guard or common input to the amplifier may be connected to a third electrode placed on the subject's forehead or attached to the arm of a high impedance potentiometer, the ends of which are connected to the left and right electrodes. The potentiometer is adjusted for minimum 60 cps noise. The amplifier used in the experiment described below contained a variable bandwidth low pass filter. The filter cutoff was set to 10 cps.

The system was calibrated by asking the subject to look at an object placed midway between the two displays. The amplifier offset was used to zero the amplifier's output. With a voltage gain of $2500 \pm 15^\circ$ of eye motion produced ± 1 volts dc. The electro-ocular potentials of some subjects contain a small dc offset in addition to the potentials due to eye motion. These offsets may be removed by placing the offset correction circuit shown in figure 5 in series with one of the electrodes. The polarity of the correction is reversed by reversing the battery. The potentiometer is adjusted periodically to insure that the amplifier output is within its specified voltage range.

A typical eye motion record is shown in figure 6. As can be seen from this figure, no difficulty is encountered in determining which display is being viewed.

THE EXPERIMENTAL DATA

As described previously, a single tracking run from a well-trained subject was digitized and stored on a bulk storage disk file. All of the results described below were derived from this digital record. A typical portion of the tracking data is seen in figure 7.

ARCHITECTURE OF THE PROPOSED HUMAN OPERATOR MODEL

The architecture of the proposed human operator model is shown in figure 8. The model may be separated into three basic systems: the scanning and information processing system which operates on the displayed signals to produce the information utilized by the operator's mental processes, the eye motion system, and the manual control system. If a structure is assumed for the scanning and information processing system, then the known eye position time history can be used to compute the operator input signals. Once these are known, the closed-loop human operator model can be opened, allowing the eye motion system and the manual control system to be treated independently.

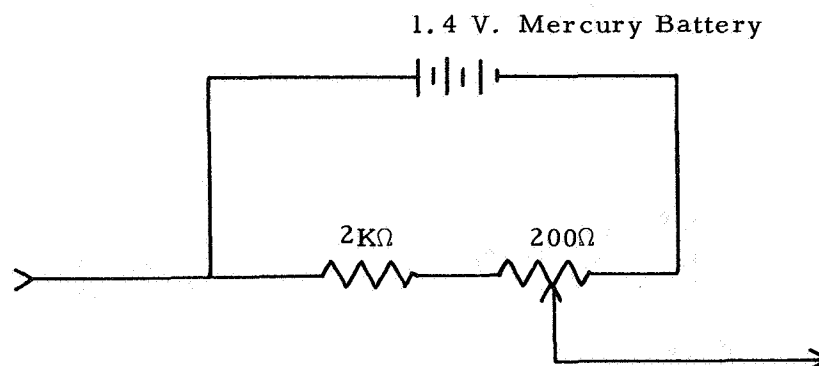


Figure 5.—dc potential correction circuit.

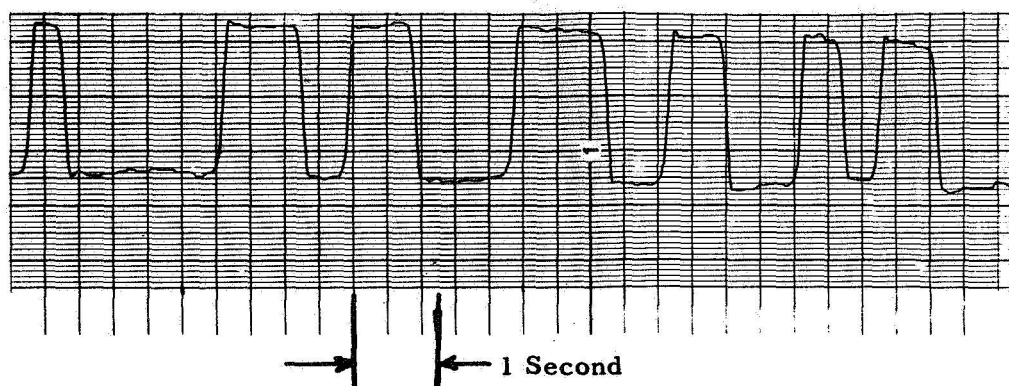


Figure 6. —A typical eye motion time history.

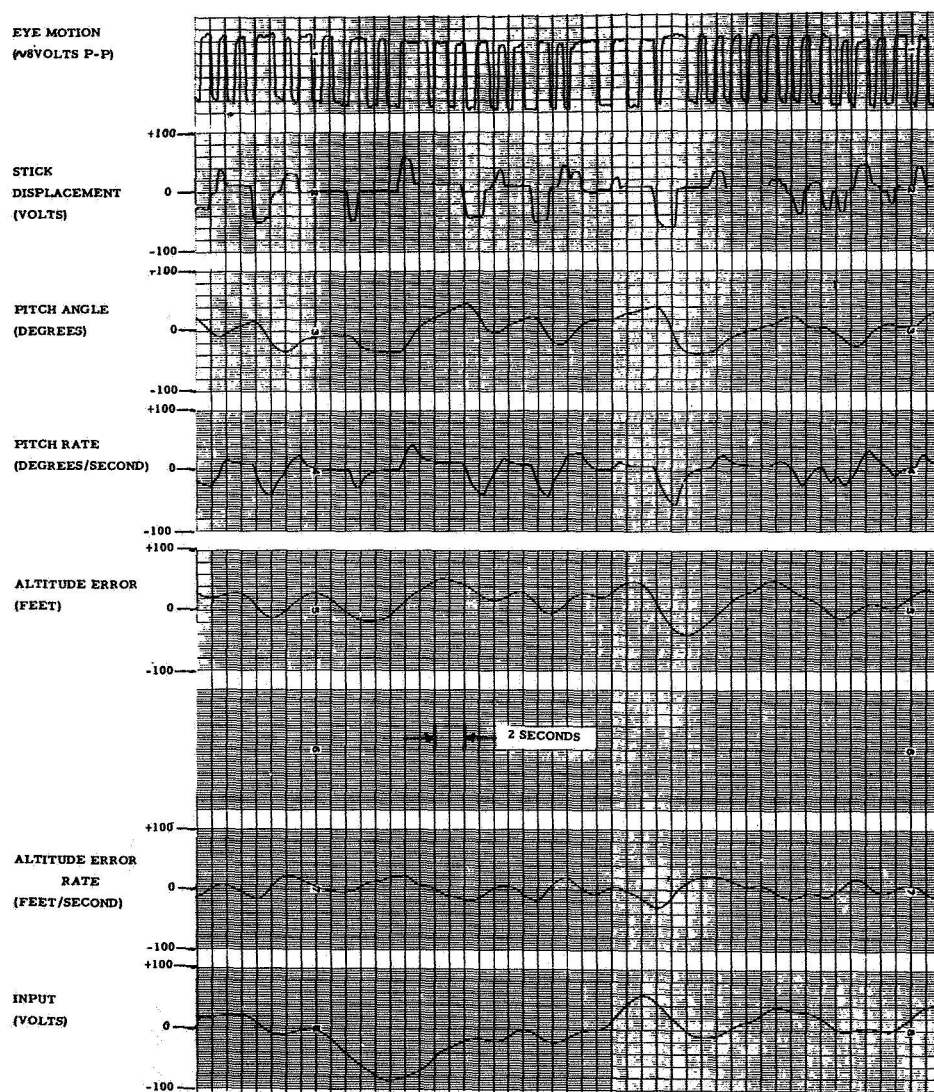


Figure 7. —A typical portion of the tracking record.

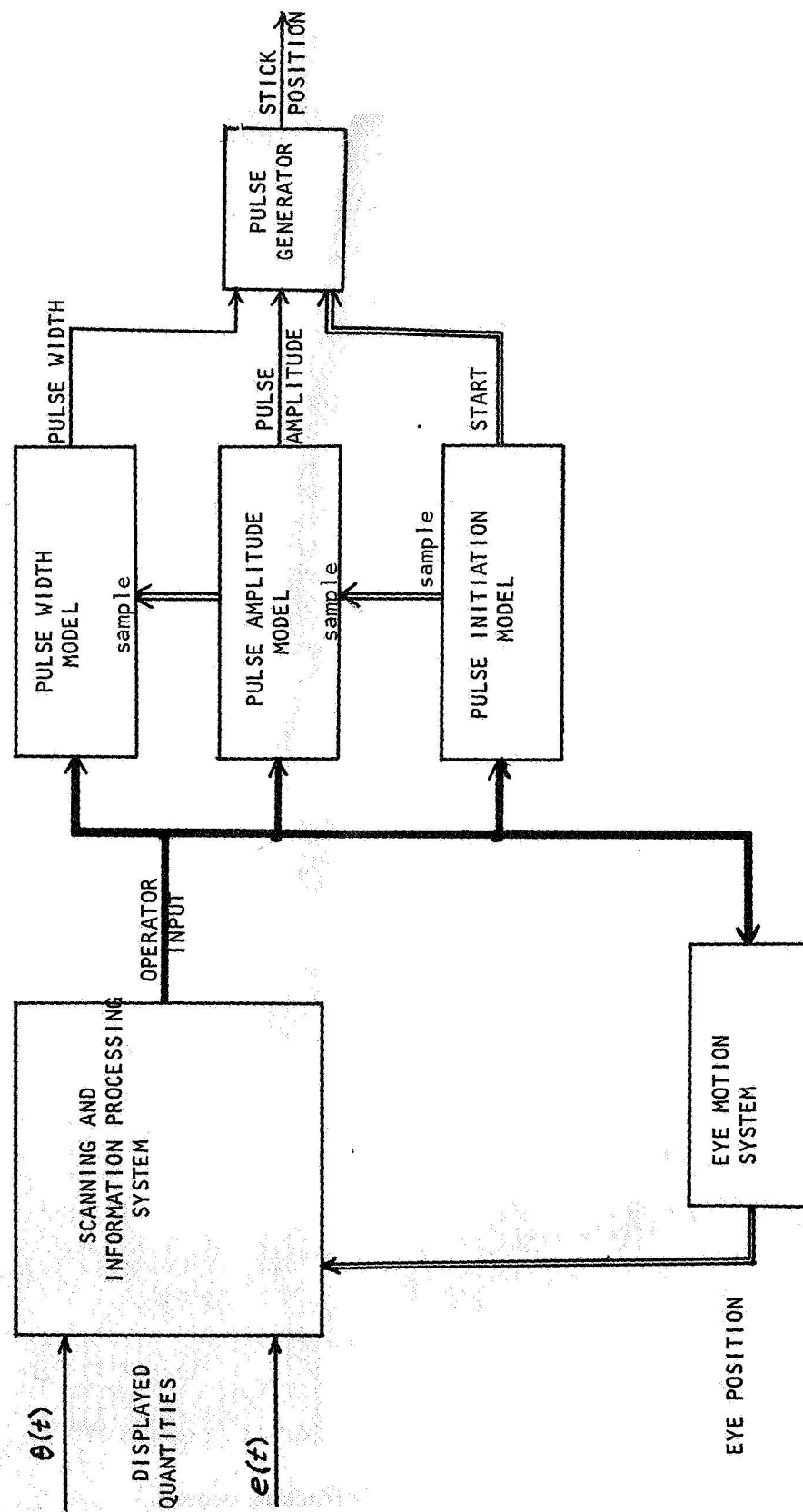


Figure 8.—Architecture of the proposed human operator model.

THE SCANNING AND INFORMATION PROCESSING SYSTEM

The scanning and information processing system cannot be identified in the usual manner as its outputs are not observable signals in the human operator. Instead, a structure is assumed, the operator input signals computed, and the remaining two systems identified.

The scanning and information processing system shown in figure 9 is based on the following hypotheses:

(1) The short fixation times and the wide separation of the displays minimizes the effects of peripheral vision.

(2) When a display is fixated upon, both the displayed signal and its derivative are available to the operator continuously.

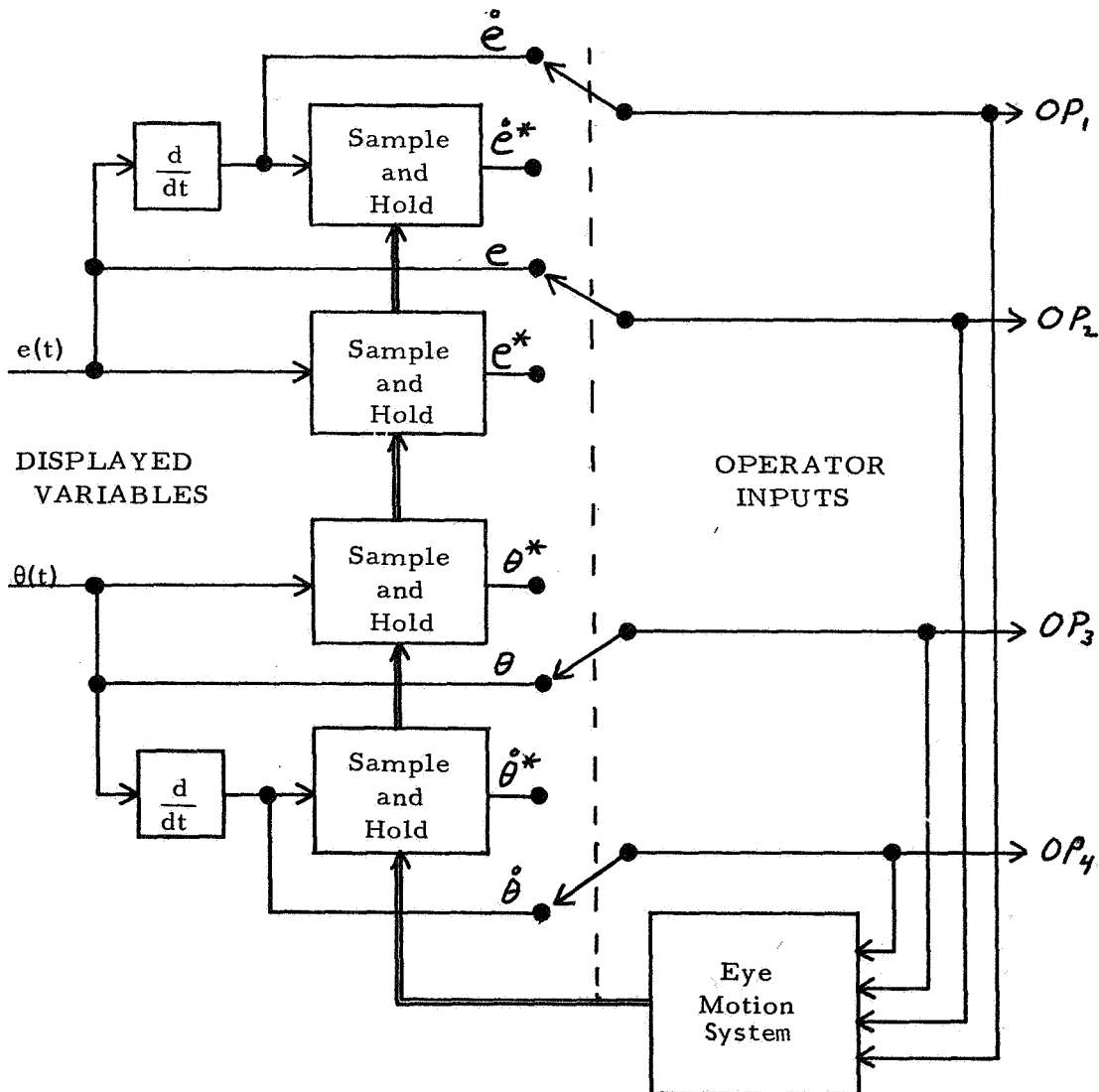


Figure 9. —Hypothesized structure of the scanning and information processing system.

(3) When a transition between displays is begun, the values of the displayed signal and its derivative are sampled and held when the eyes reach the midpoint between displays. From the midpoint on, only the sampled and held signals are available to the operator.

Figure 9 should be viewed as a first-order model to study the feasibility of the techniques described below. The role of peripheral vision, the rate at which operators estimate derivatives of displayed quantities, the exact nature of the information the operator remembers from the previous display, how long the information is used before it is obsolete, etc., are all questions that must be studied before the models can be used to make inferences concerning the cerebral processes of human operators.

THE MANUAL CONTROL SYSTEM

The manual control actions observed in figure 7 are strongly pulsatile. Two previous studies (refs. 7 and 8) have dealt with the synthesis and identification of models for pulsatile control actions using one and two displays. The manual control system for the task described in this paper is presented in reference 7 and will not be treated further here.

ON THE EYE MOTION DATA

The experimental task was designed to require eye motion commutation between two separated displays. The commutation consisted of fixations on the displays separated by rapid transitions between displays. The eye motion time history was divided into four phases:

- (1) fixated on the left display
- (2) fixated on the right display
- (3) moving from left to right
- (4) moving from right to left

The eye motion time history was found to contain 216 eye motion cycles, from left to right and back again. A digital computer program was written to determine the times, relative to the start of the run, of the four phases of each cycle. This data was used to compute the length of time spent in each of the phases, from which the distribution functions of the length of time for each phase was computed. A substantial amount of insight into the operator's eye motion control policy can be gained from an examination of these distribution functions, figures 10 and 11. The distribution functions of the two transition times are practically identical. The time spent in transition between displays ranged from 50 to 200 milliseconds, with a mean of 125 milliseconds. Thus the time of a transition may be assumed to be independent of the direction of the transition.

The fixation interval distribution functions both have a peak at approximately 600 milliseconds. However, the attitude display (left) fixation times range from 200 milliseconds to as much as 2 seconds, with a large number of fixations longer than 1 second. The altitude error (right) display is rarely fixated on for more than 1 second. The distribution function for the right fixation times exhibits a strong, narrow peak at 600 milliseconds. A similarly shaped distribution function would have occurred if the operator had sampled the right display with a mean sampling interval of 600 milliseconds. If this were indeed the case, then the operator's visual scanning policy is to concentrate on the attitude (left) display until such time as it is necessary to sample the altitude error (right) display.

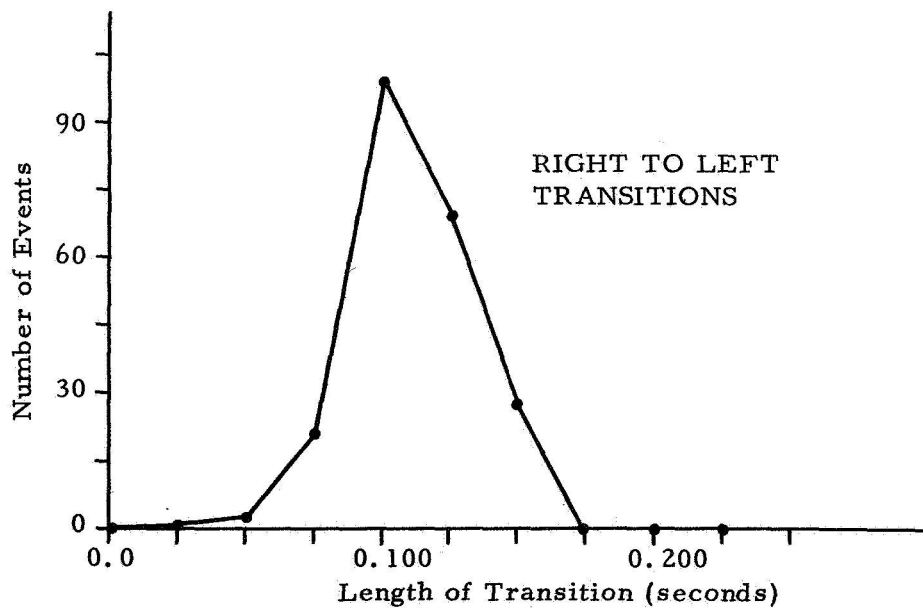
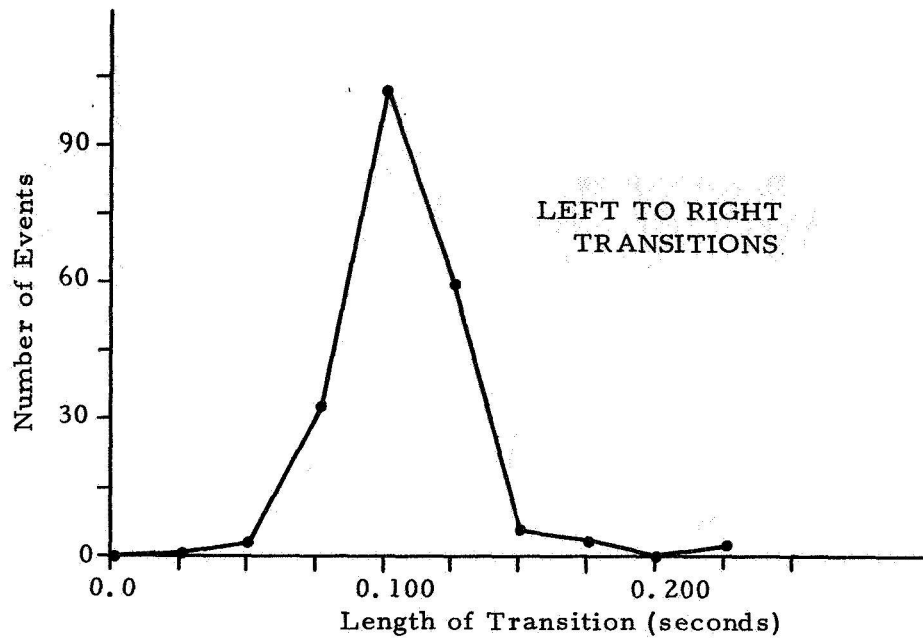


Figure 10.—Distribution functions of eye motion transitions.

THE EYE MOTION SYSTEM VIEWED AS A MULTISTATE DECISION PROCESS

The input to the eye motion system consists of four signals, two continuous and two sampled and held. These signals may be viewed as a four-dimensional input space, and the decision process as one or more hypersurfaces that divide the input space into left fixation regions and right fixation regions. When a hypersurface is crossed from one region to another an eye motion transition results. The multistate decision element and its associated

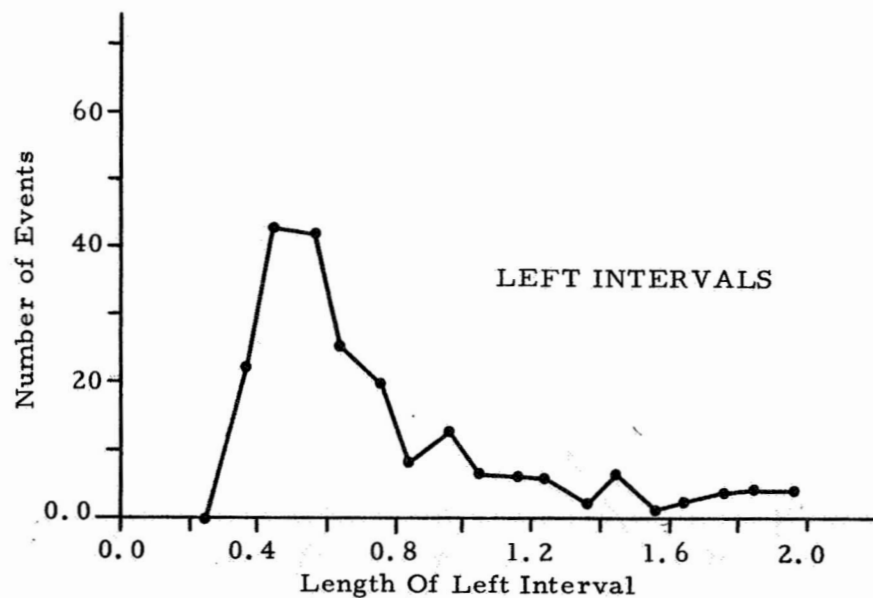
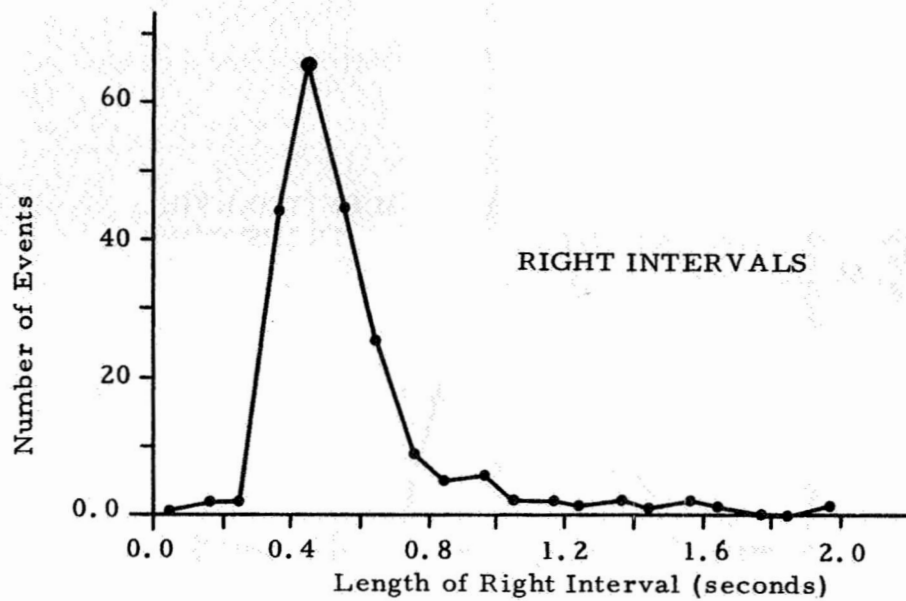


Figure 11.—Distribution function of eye fixation intervals.

identification algorithm, described in reference 7, was designed to model such a decision process. An MSDE, which mechanizes these concepts, is seen in figure 12. Three hyperplanes are generated in the input space. The threshold elements determine whether the point (OP_1 , OP_2 , OP_3 , and OP_4) is above or below the hyperplane. The outputs of the threshold elements are gated together to generate the desired output states. The time-delay element may be included if it is desired to simulate the computation time or delay associated with the human operator's mental processes and neuromuscular system. The generation of more complex surfaces in the OP_1 , OP_2 , OP_3 , and OP_4 space may be viewed in two ways:

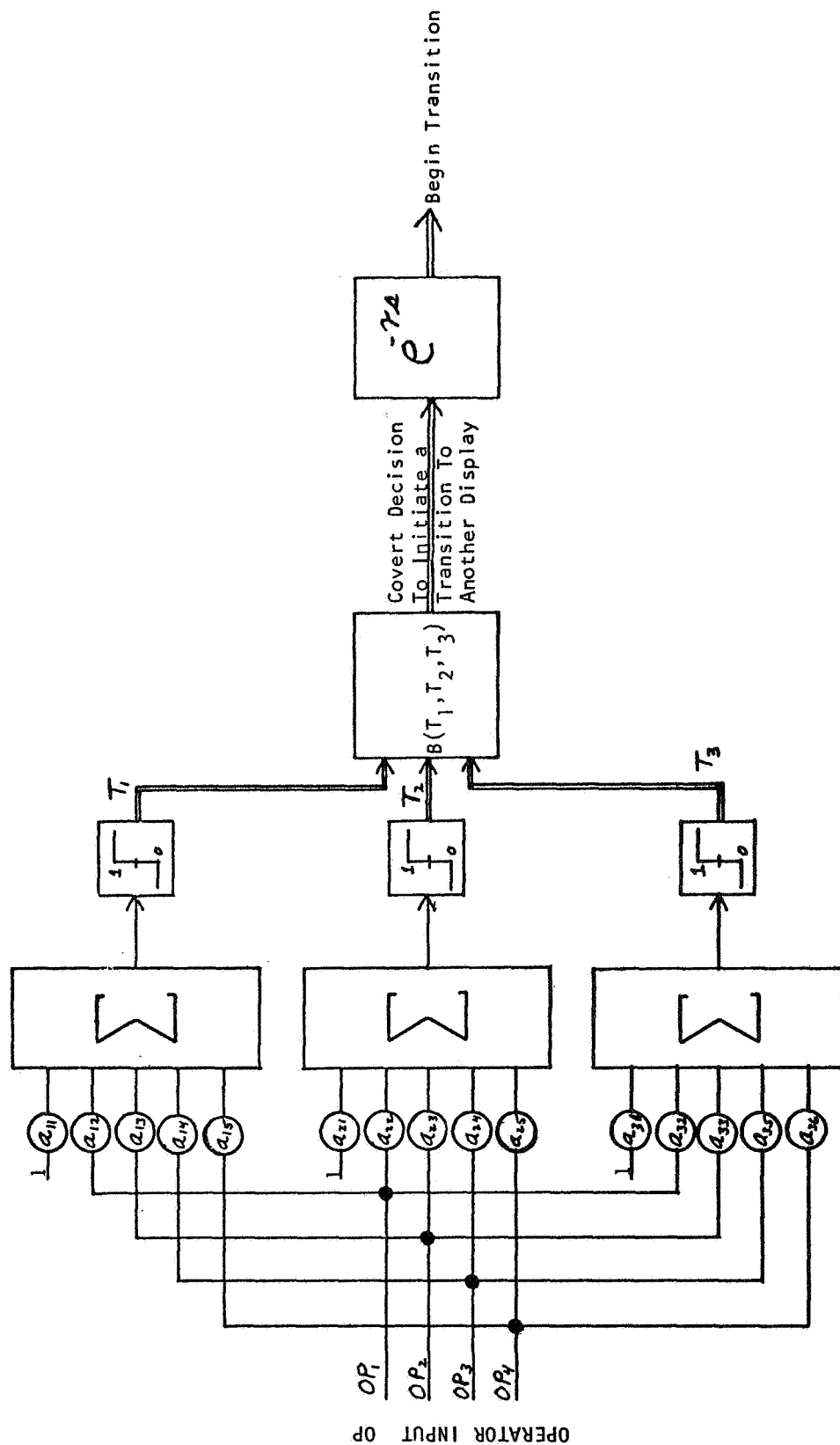


Figure 12.—Eye motion as a multistate decision process.

(1) Add additional input terms to the input space consisting of squares, cubes, and cross products of the four operator input signals. The resultant surfaces are hyperplanes in a higher dimensional space. (2) View the Boolean combination of the threshold outputs as forming a four-dimensional function generator that utilizes hyperplanes to approximate more complex surfaces.

The first viewpoint lacks generality as it requires a priori knowledge concerning the number and type of nonlinear functions of the operator input signals to include. The later viewpoint requires no assumptions; additional hyperplanes are added, and the Boolean function changed until a satisfactory model is achieved.

The MSDE is identified by selecting a set of starting coefficients, a_{ij} , and a Boolean function $B(T_1, T_2, \text{ and } T_3)$, usually n out of 3. The OP_1 — OP_4 time histories just prior to the termination of a left fixation interval are applied to the input of the MSDE and the coefficients and n adjusted until the MSDE generates an output τ seconds prior to the actual termination of the fixation. If the MSDE fails to generate an output, a missed event is recorded, and the process resumed at the conclusion of the next right fixation.

One thousand sets of the coefficients a_{ij} through a_{35} were selected from computer generated uniform random numbers for each value of $n=1, 2, \text{ or } 3$. The best 10 sets for each n were used as starting conditions for the optimization algorithm. In all cases the MSDE's produced 80 percent or more of their outputs as soon as the operator input signals were applied. It is impossible to determine whether this failure resulted from the nature of the process being modeled or a failure of the experimenter to select the correct starting values for the coefficients.

The dimensionality of the input data precludes the use of intuition and insight to select either starting values for the coefficients or even the basic shape the decision surfaces might take. If only two inputs, or at most three inputs were involved, a digital computer could be used to sketch the input space trajectories just prior to the termination of the left fixation intervals. This would allow the experimenter to select approximate decision hyperplanes.

THE EYE MOTION SYSTEM VIEWED AS A PROPORTIONAL DECISION PROCESS

An alternate hypothesis concerning the operator's visual scanning mechanism is that the length of the fixation interval is functionally related to the operator input signals observed during the transition toward the display. The proportional decision element (PDE) shown in figure 13 was developed to model such a process. The PDE shown in figure 13 is easily expanded to describe nonlinear relationships between the lengths of the fixations and the operator input signals. The PDE actually used to model the eye motion data was supplied with the following nonlinear functions of the operator input:

$$\begin{aligned} OP_1 &\approx \theta(t) \\ OP_2 &\approx \theta(t) \\ OP_3 &\approx e(t) \\ OP_4 &\approx \dot{e}(t) \end{aligned}$$

t_i = time at the beginning of the i th fixation on the right display

$$\begin{aligned} x_0 &= 1.0 \\ x_1 &= OP_1 = \theta(t_i) \end{aligned}$$

FUNCTIONS OF THE OPERATOR INPUT OP_1 , OP_2 , OP_3 , OP_4

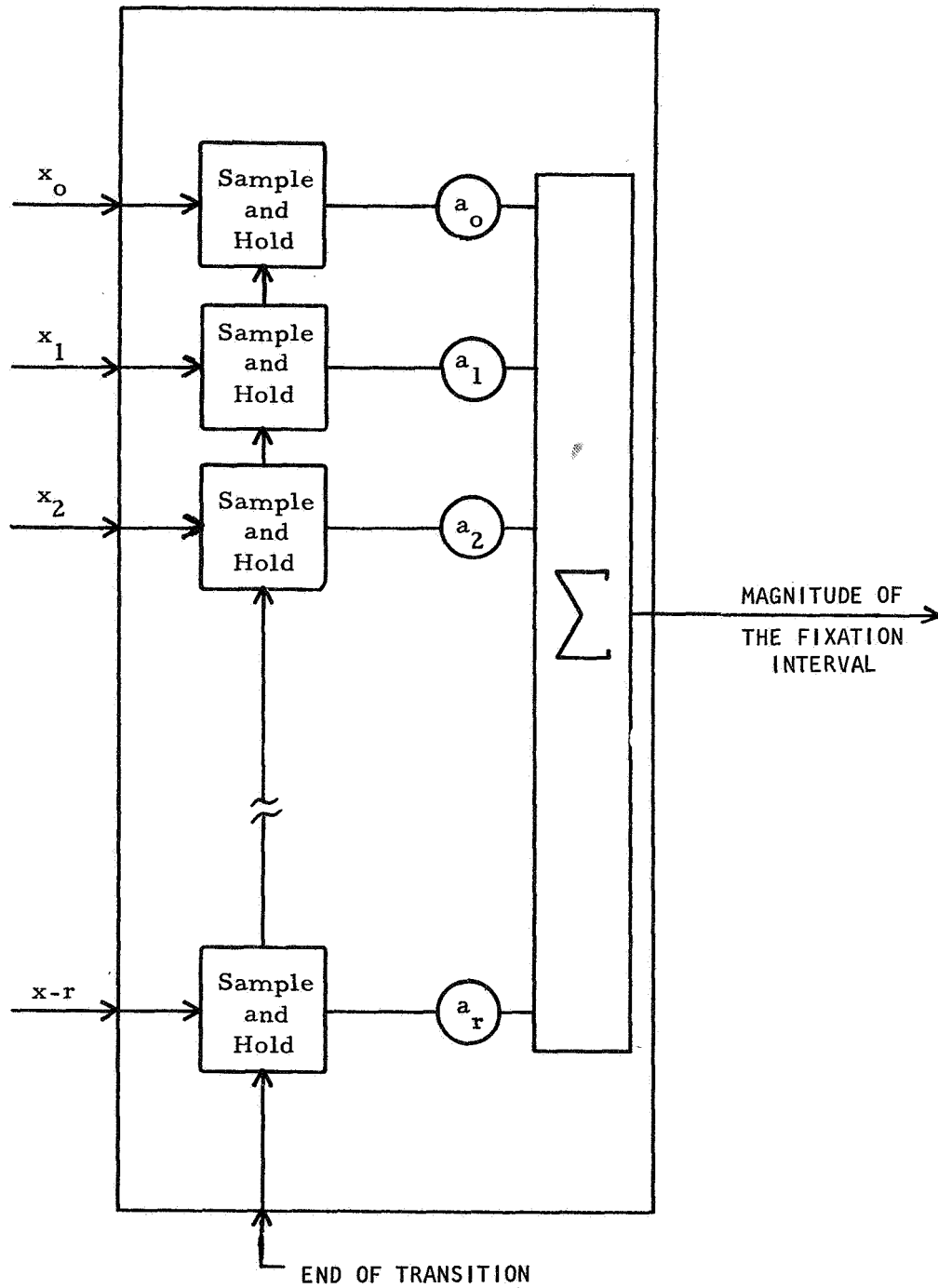


Figure 13. —A proportional decision element eye motion model.

$$\begin{array}{lll}
 x_2 & = OP_2 & = \theta(t_i) \\
 x_3 & = OP_3 & = e(t_i) \\
 x_4 & = OP_4 & = \dot{e}(t_i)
 \end{array}$$

$$\begin{aligned}
x_5 &= OP_1^2 &= \theta^2(t_i) \\
x_6 &= OP_2^2 &= \theta^2(t_i) \\
x_7 &= OP_3^2 &= e^2(t_i) \\
x_8 &= OP_4^2 &= \dot{e}^2(t_i) \\
x_9 &= OP_1 * OP_2 &= \theta(t_i) * \theta(t_i) \\
x_{10} &= OP_1 * OP_3 &= \theta(t_i) * e(t_i) \\
x_{11} &= OP_1 * OP_4 &= \theta(t_i) * \dot{e}(t_i) \\
x_{12} &= OP_2 * OP_3 &= \theta(t_i) * e(t_i) \\
x_{13} &= OP_2 * OP_4 &= \theta(t_i) * \dot{e}(t_i) \\
x_{14} &= OP_3 * OP_4 &= e(t_i) * \dot{e}(t_i)
\end{aligned}$$

The inputs supplied to the PDE correspond to the terms of a four-dimensional quadratic surface in the OP space. The output of the PDE is given by:

$$f_i = a'x_i$$

where f_i is the PDE output, x_i is an n vector of sampled and held signals, and a is an n vector of unknown coefficients. The performance of the model is measured by:

$$\phi = \frac{1}{216} \sum_{i=1}^{216} (f_i - l_i)^2$$

where l_i is the actual length of the i th fixation interval. The identification procedure consists of selecting the coefficient vector, a , such that ϕ , the mean squared model error, is minimized. The complete identification procedure and the resultant coefficients are given in reference 7. The fidelity of the model may be studied by plotting the model output versus the actual quantity. If the model is perfect and the data undistorted by noise, then all of the data points lie on a unity slope straight line. The resultant plot for the right fixation intervals is seen in figure 14. The model exhibits good correlation with the experimental data for fixation intervals up to 900 milliseconds. For fixation intervals longer than this, the model deviates considerably from the experimental data. Attempts to remedy the situation by increasing the PDE to a cubic surface or by adding inverse terms produced no significant improvements in the model responses.

A DUAL PDE, MSDE EYE MOTION MODEL

The model developed in the preceding paragraph accurately describes short right fixation intervals. It was hypothesized that the operator utilized different decision processes depending on the magnitudes of the operator input signals. In such a case, an MSDE could be used to partition the input space. The data points in each partition would be treated as separate modeling problems.

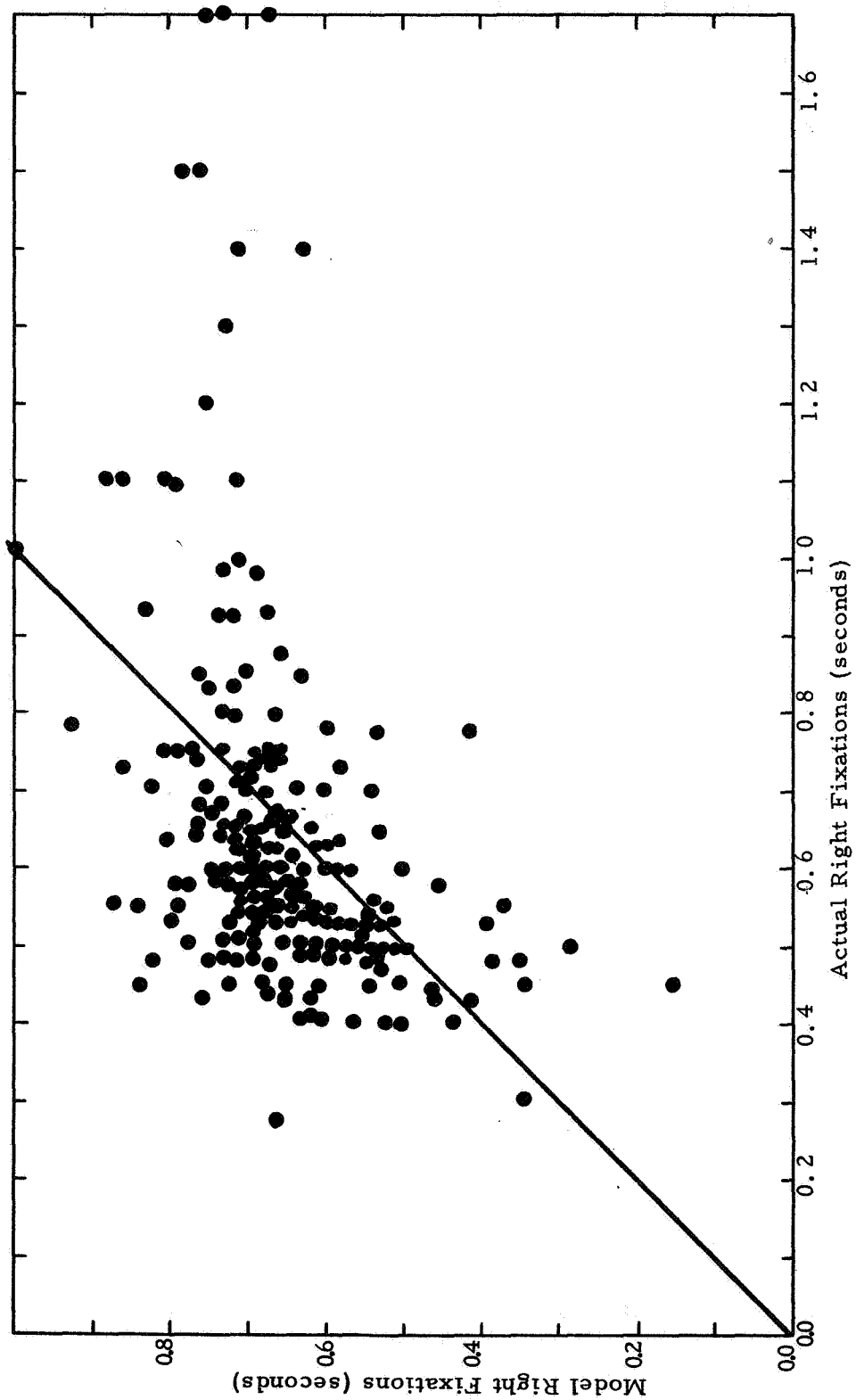


Figure 14. -- Length of PDE right fixation intervals versus actual right fixation intervals.

An examination of the tracking data revealed that leisurely tracking occurs in portions of the phase space where the generation of a pulsatile stick motion is unlikely. This occurs when:

- (1) $e(t)$ and $\dot{e}(t)$ are opposite in sign and $\dot{\theta}$ is large
- (2) all four signals are small.

The MSDE seen in figure 15 partitions the phase space accordingly. The MSDE's output is described by:

$$\bar{b} = \begin{cases} 1 & \text{if } e, \dot{e} \text{ are opposite in sign and } \dot{\theta} > r_1 \\ & \text{or } |e| + |\dot{e}| + |\theta| + |\dot{\theta}| > r_2 \\ 0 & \text{if otherwise} \end{cases}$$

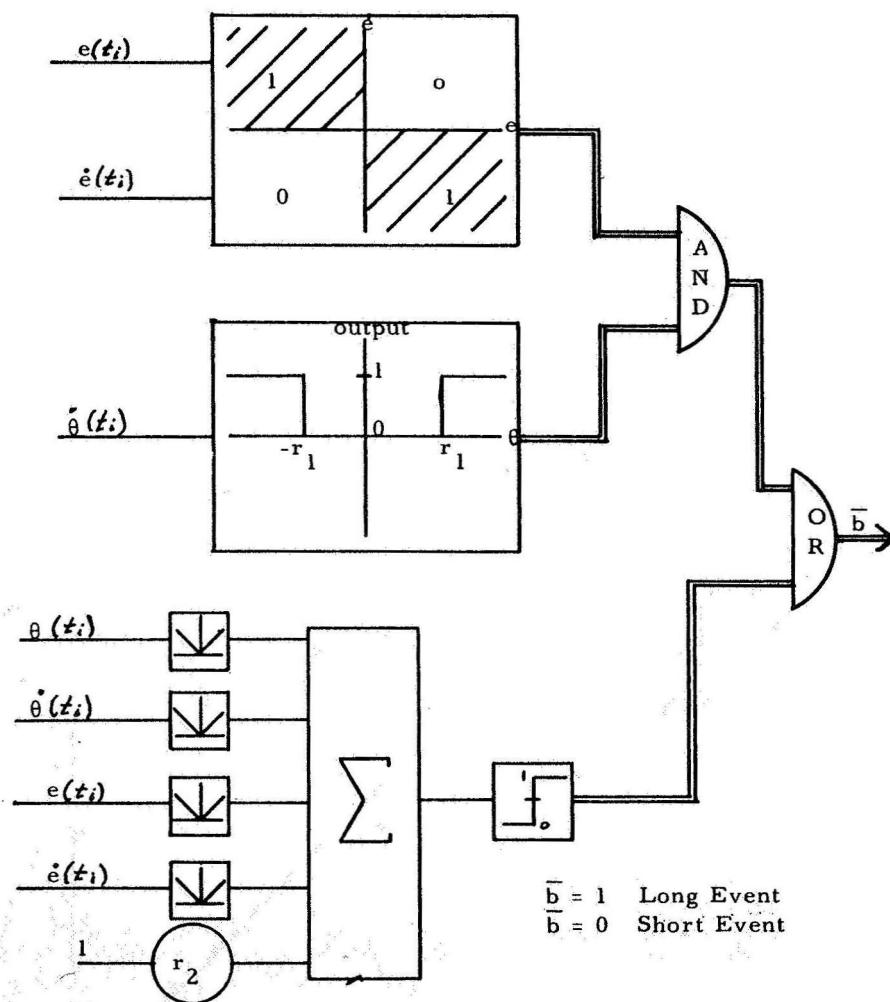


Figure 15. —Use of an MSDE to partition the input space for a dual PDE eye motion model.

For fixed values of r_1 and r_2 , the data points x_i can be partitioned into long event points and short event points, and the two groups of data points modeled independently. The resultant model structure is seen in figure 16. A systematic study of values of r_1 and r_2 was carried out. The results for the optimal values of r_1 and r_2 are seen in figures 17 and 18. With the exception of the four data points seen in figure 17, the dual PDE, MSDE model is quite satisfactory.

The selection of the MSDE decision surfaces was based on intuition and two-dimensional pictures of a four-dimensional space. The purposes of this study are satisfied by finding one MSDE. The study of more complex partitionings of the input space is clearly a necessary extension of this work.

THE COMPLETE EYE MOTION MODEL

If it is assumed that a similar modeling process will suffice to describe the left fixation data, then the addition of a storage element and timing devices completes the eye motion model as shown in figure 19.

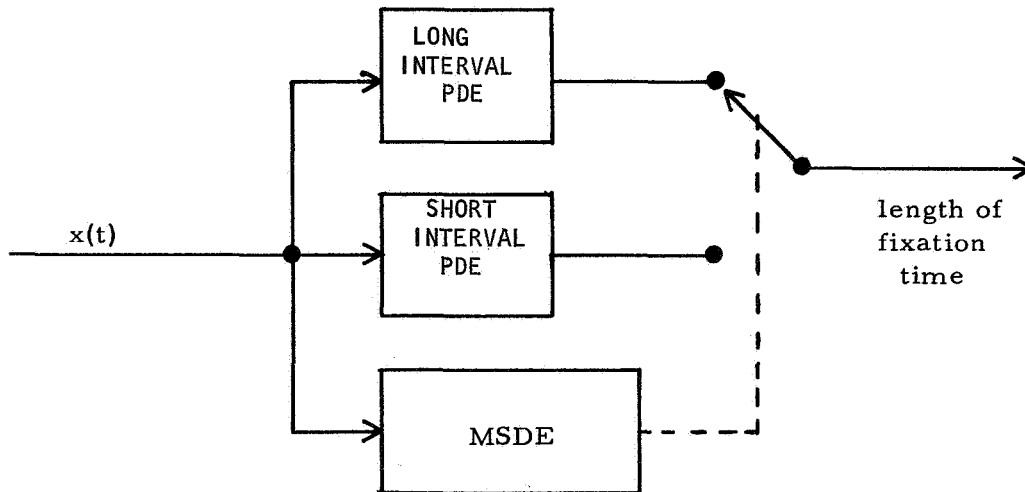


Figure 16.—Dual PDE MSDE eye motion model.

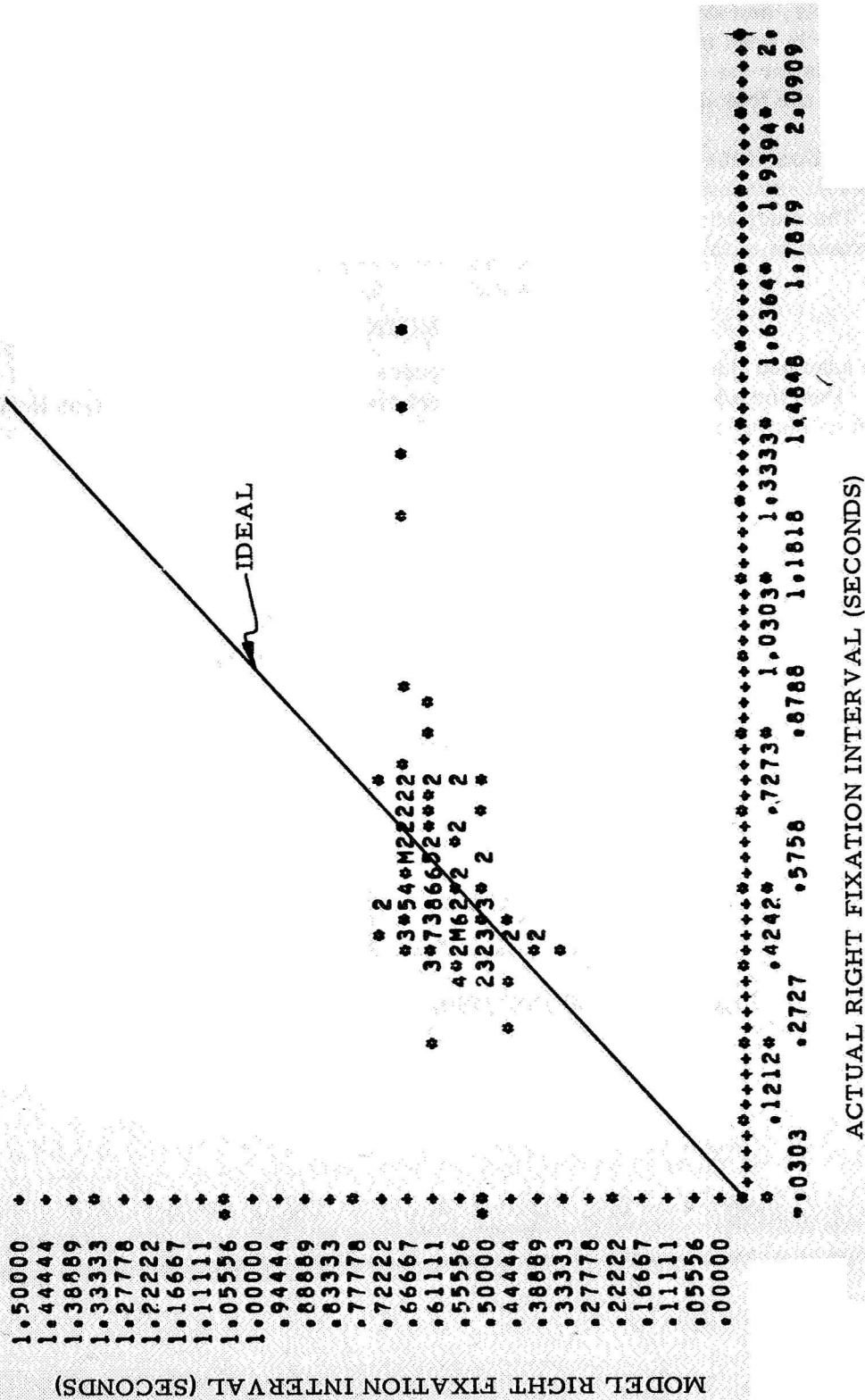


Figure 17. —Model fixation length versus actual fixation length for right display-short PDE.

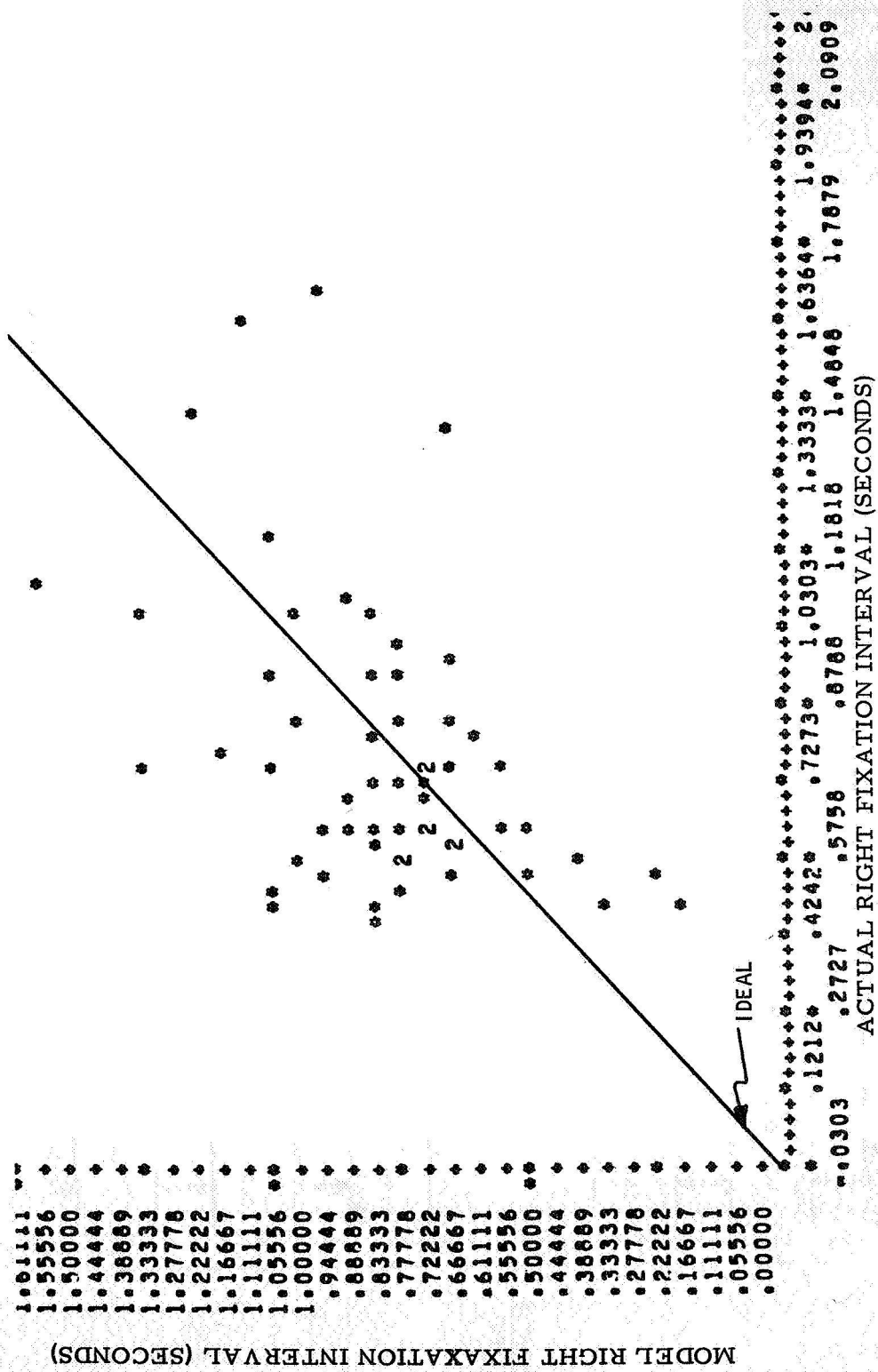


Figure 18. — Model fixation length versus actual fixation length for right display-long PDE.

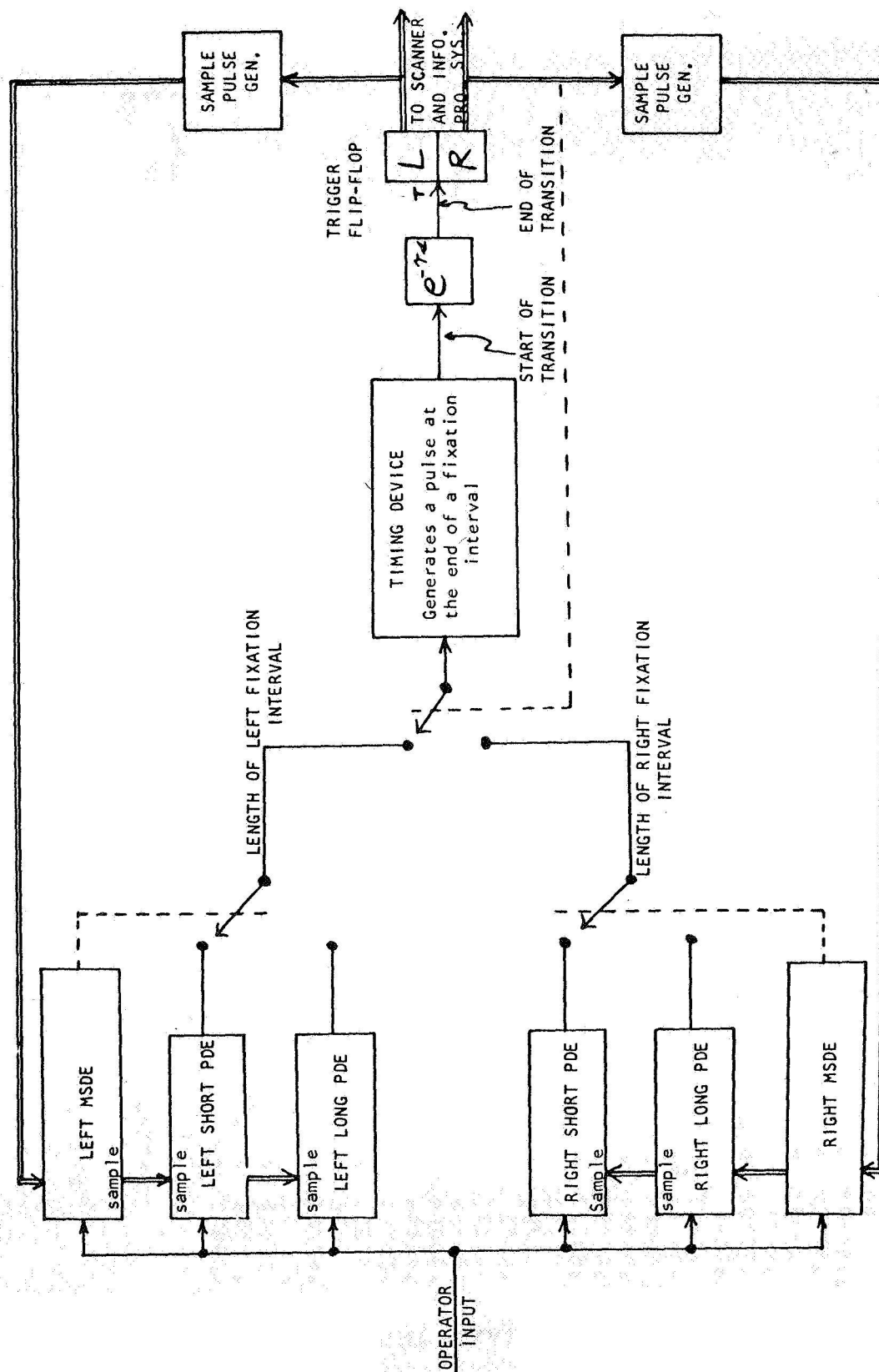


Figure 19. —Complete eye motion model.

REFERENCES

1. Senders, J. W.: The Human Operator as a Monitor and Controller of Multi-Degree of Freedom Systems. 4th National Symposium on Human Factors in Electronics, Washington, D. C., May 1963.
2. Senders, J. W.; and Stevens, K. W.: A Re-Analysis of the Pilot Eye Movement Data. Bolt, Beranek and Newman, Inc., Rept. No. 1136, May 1964.
3. Senders, J. W.; Elkind, J. I.; and Smallwood, R.: An Investigation of the Visual Sampling Behavior of Human Observers. Bolt, Beranek, and Newman, Inc., Rept. No. 1246, May 1965.
4. Carbonell, Jaime R.: A Queueing Model of Many-Instrument Visual Sampling. IEEE Trans. Human Factors in Electronics, vol. HFE-7, no. 4, Dec. 1966.
5. Levison, W. H.; and Elkind, J. I.: Two-Dimensional Manual Control Systems with Separated Displays. USC-MIT Working Conference on Manual Control, Mar. 1967.
6. Merritt, M. J.: An Asynchronous Pulse Amplitude-Pulse Width Model of the Human Operator. NASA SP-144, 1967, pp. 225-240.
7. Merritt, M. J.: Synthesis and Identification of Mathematical Models for the Discrete Control Behavior of Human Operators. Univ. Southern California, USCEE Rept. No. 202, May 1967.
8. Young, L. R.: Methods of Oculography. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Mass.
9. Schott, E.: Uber die Registrierung des Nystagmus und anderer Augenbewegungen vermittels des Saitengalvanometers Deutsches Arch. Klin. Med., vol. 140, 1922.
10. Mowrer, O. H.; Ruch, R. C.; and Miller, N. E.: The Corneo-Retinal Potential Difference as the Basis of the Galvanometric Method of Recording Eye Movements. Am. J. Physiol., vol. 114, 1936.
11. Pew, Richard W.: Performance of Human Operators in a Three State Relay Control System with Velocity-Augmented Displays. IEEE Trans. Human Factors in Electronics, vol. 7, no. 2, June 1966.
12. Pew, R. W.; Duffendock, J. C.; and Fensch, L. K.: On the Relation Between Discrete and Continuous Performance. 7th IEEE Symp. Human Factors in Electronics, May 1966.

IV. ANALYSIS OF ADAPTIVE OR HIGHER ORDER CONTROL PROCESSES

17. Stochastic Modeling of Human Learning Behavior

Jacob L. Meiry
Massachusetts Institute of Technology.

The application of a stochastic model of human learning behavior in manual control tasks is extended to the performance of the human operator in regulating all second-order dynamic processes. In particular, human control decisions for optimal multiswitch and unstable dynamic systems are found to correlate well with responses recorded by the model in the study of simulated subjects. The universality of the model is discussed in the light of a series of compensatory task experiments.

INTRODUCTION

A theory of human learning behavior in the regulation of a second-order dynamic process by actuating a two position controller was previously presented in a paper entitled "Stochastic Modeling of Human Learning Behavior " in reference 1. The concept of skill organization is outlined as a statistical revision process applied to the identification of a control policy for the manual control task. To characterize the decisionmaking of the human operator, we have postulated a single-channel information system for the subject. A hypothetical decision center is entrusted with selecting among response alternatives at the moment of choice. The theory represents individual preferences for control policies as probabilities, continuously revised on the basis of experimental outcomes. Specifically, our interpretation of Bayesian statistics is the principal factor underlying the process of human decisionmaking.

The model of human learning behavior was submitted as a uniform control strategy applicable to all second-order, first-degree dynamic processes. Its success in representing the time-variant random nature of human learning in a series of controlled experiments was unqualified. This paper reports a new sequel of experiments to verify our theory and model operation when called to control harmonic oscillators and unstable plant dynamics. Finally, learning of a compensatory task is compared with human operator performance in controlling a pure inertia plant to minimize the integrated absolute error.

THEORY REVIEW

The stochastic information system postulated for the human operator consists of a sensor, a decision center, and an effector, all serially connected. The sensor can perceive quantized state information since we assume a finite grid to overlay the state space of the

dynamic process. The decision center in turn processes the state information and decides upon a response. For the ideal relay controller this involves a decision: whether or not control polarity should be reversed. To visualize the process, the concept of a switch curve is helpful. By definition, the switch curve is the locus of all points in the state plane where a control polarity reversal should occur to force the second-order process' trajectory through the null state. As a matter of control policy, the human operator has to decide whether the state of the process has reached the location of the switch curve. We have proposed the use of subjective probabilities for this identification procedure. Thus we assume that stored in the decision center memory are columns of probabilities expressing the subjective estimate of the mesh where the switch curve lies, with higher probability reflecting higher confidence in the location of the curve. The strategy of the control policy was made dependent upon a switching probability evaluating the belief that the switch curve lies beneath the present state of the process. Indeed, if the switching probability is high, one is quite confident that a switch should take place. Selection of alternatives, then, can be based on the value of this probability. However, to account for variations in human control behavior, it is assumed that the decision to switch is a Bernoulli trial with the switching probability being the probability of success.

It is quite evident that learning a psychomotor task, then, is a revision of opinion about the location of the switch curve in the state space. In terms of probabilities, a weighting of previous decisions could be used to improve one's concepts about the appropriate strategy to follow. This is weighting of evidence, used here to evaluate the significance of an outcome of a decision in relation to the expected outcome. The revision of prior probabilities was assumed to constitute an application of Bayesian statistics, although this efficient algorithm exaggerates the decisionmaking capabilities of the human. Our theory as presented previously examines the information used as evidence and the formulation of posterior probabilities for future trials. The reader is referred to the above-mentioned paper for the mathematical treatment of the theory and the model.

THE EXPERIMENT

Three groups of eight subjects at MIT's man-vehicle laboratory were given the opportunity to learn a manual control task. They were briefed on the task and familiarized with the apparatus but were not allowed to practice prior to the first trial. The subject was required to null the initial misalignment between two line segments displayed on an oscilloscope by actuating a two position switch. The left segment remained stationary and the displacement of the other relative to it, x , satisfied a second-order differential equation. Each subject was given 50 5-second trials starting with the same initial conditions at 10-second intervals. Subject performance on each trial was measured by computing the integral of the absolute value of x over the 5 seconds, and this score was reported to the subject immediately after each trial. The dynamic processes controlled were of the form

$$\begin{aligned}\ddot{x} + x &= u & u &= 4 \text{ cm/sec}^2 \\ \ddot{x} + 4x &= u & u &= 7.5 \text{ cm/sec}^2 \\ \ddot{x} - x &= u & u &= 7.5 \text{ cm/sec}^2\end{aligned}$$

to represent the range of distinct response characteristics of second-order dynamics. A fourth group was engaged in learning to control a compensatory tracking task with the sum of 10 sine waves as input. The input signal is in accordance with that of reference 2. The process dynamics were pure inertia ($\ddot{x} = u$).

RESULTS AND DISCUSSION

The first series of experiments dealing with a regulator task is quite evidently designed to answer the following questions:

- (1) Are model and subjects learning the optimal control strategy for difficult dynamic processes?
- (2) Are model and subject responses drawn from the same parent population?
- (3) Is a unique model of control strategy representative of decisionmaking in control of second-order dynamic processes?

To underline the significance of these experiments in evaluating the theory of human learning behavior, let us briefly discuss the plant dynamics selected. Since the optimal system response (with respect to the minimum integrated absolute error) is practically identical to the time optimal system (ref. 3), an easy comparison of control quality is feasible if time optimal control is used as a reference. Unlike the pure inertia dynamic system, the harmonic oscillator has imaginary eigenvalues. The time optimal control of the oscillator can involve multiple switching unlike a real-root eigenvalue system. Moreover, the optimal trajectories are a series of semicircles forcing the human operator to assume different control strategies for different parameters of the oscillator. One would expect the subjects to adopt a suboptimal policy whenever more than one switch is required. Thus the simulated dynamics of two oscillators were examined with expectations of finding a suboptimal control behavior for the oscillator with $\omega = 2$ rad/sec. On the other hand, the unstable plant is characterized by a switch curve with a very large gradient, while the harmonic oscillator has a relatively flat switch curve. Quite obviously the learning potential of the model and the subjects are heavily taxed with the dynamic processes selected, offering an opportunity for strict examination of the validity of the human learning behavior model.

EXPERIMENTAL RESULTS

We chose to express the subjects' learning performance by location of the switching trajectory in the phase plane. Figure 1 represents the average location of the first switching point measured for eight subjects at their 50th trial. Similarly the switch curve of the model is obtained from the table of posterior probabilities for eight simulated subjects. Figures 2, 3, and 4 are the average learning curves in terms of the integrated absolute error (IAE) for subjects and model. While all learning curves converge to an asymptotic value, some differences are noted between respective curves for subjects and model. We submit that these differences are not significant statistically, and we support this finding with results from the Mann-Whitney U test (ref. 4) performed on the interresponse time (time between two consecutive responses). At the 1 percent level, our tests accept the hypothesis that the subject sample and the model sample are drawn from identical distributions for 95 percent of the cases we examined in this manner (ref. 5).

Our statistical tests supporting the assumption of statistical images for a sample of behavioral simulations obtained with the model and from a series of psychomotor experiments offer us a quantitative basis to conclude that the proposed theory passes the test of generality. Similarly, it is quite evident (see fig. 1) that the skilled control policies of subjects and model closely resemble each other and are executed in a nearly time optimal fashion. All factors considered, we accept the theory of human learning behavior as an explanation of human control behavior in the regulation of second-order dynamic processes simulated as manual control tasks.

The application of the theory to compensatory tracking tasks is not an obvious extension of the class of problems just treated. First, it is evident that for such tasks the evidence for the revision process is heavily influenced by the dynamic characteristics of the input. Furthermore, an ideal two level relay controller for a manual compensatory task is quite unrealistic as a design from the point of view of fuel management. On the other hand, a study of model behavior in compensatory tasks is indicative of possible human difficulties in the search process for correct control policy. We submit that the original model of human learning behavior is capable of identifying a region of switching in the phase plane. After 20 trials, this region will contain over 90 percent of the switch reversals of the human operator in the compensatory task. Although these findings might support a speculation about the universality of the model as a representation of the time varying random performance of the human operator in manual control tasks, we surmise that only the approach of the theory is general. In this context, we have offered sufficient evidence drawn from a variety of tasks to support a contention about accepting the theory's explanation of human learning.

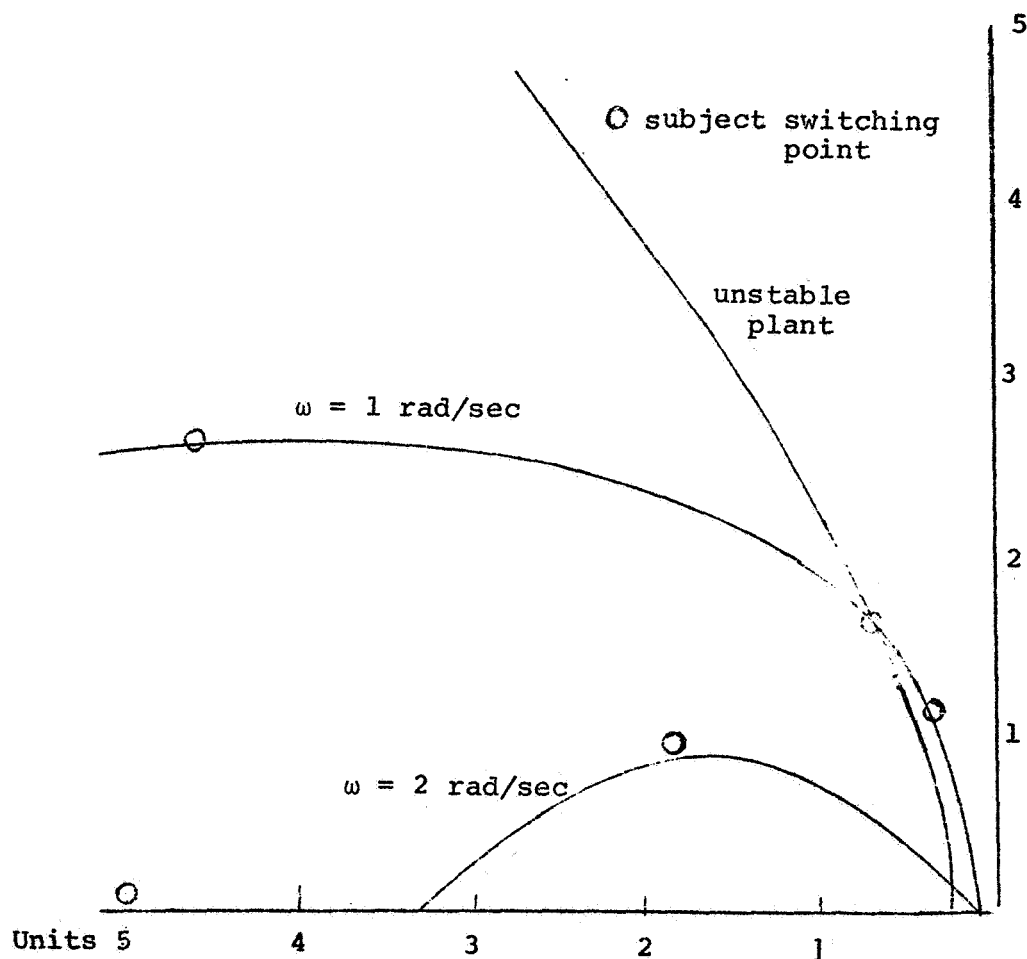


Figure 1.—Decision switching curves.

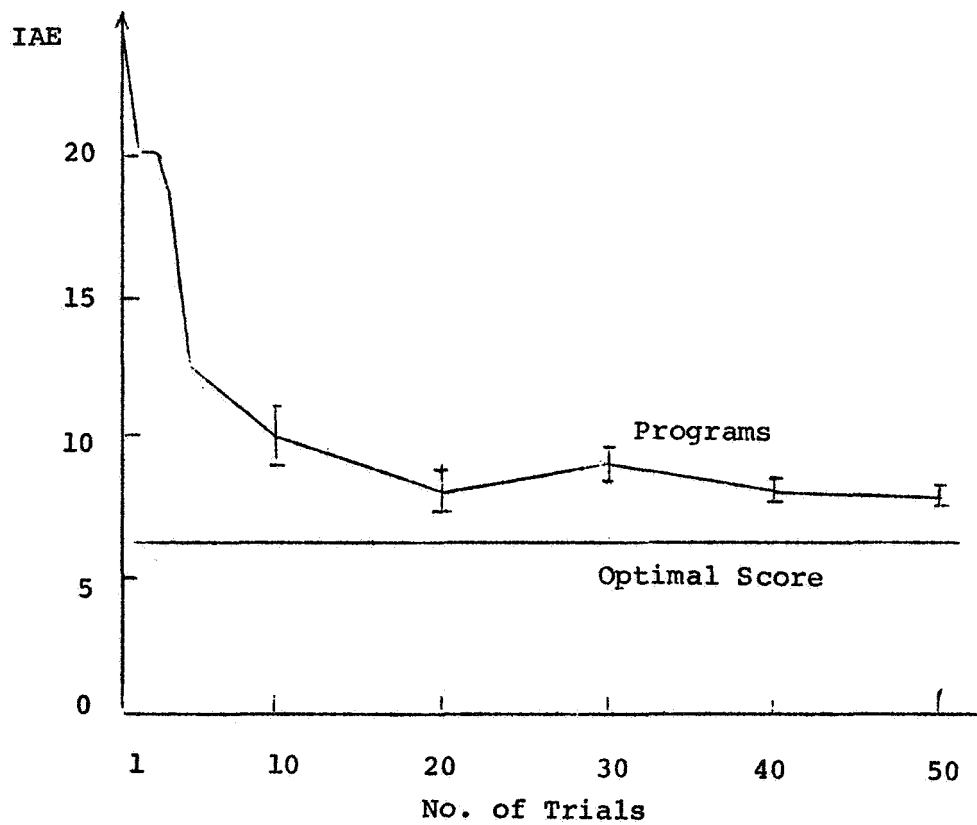
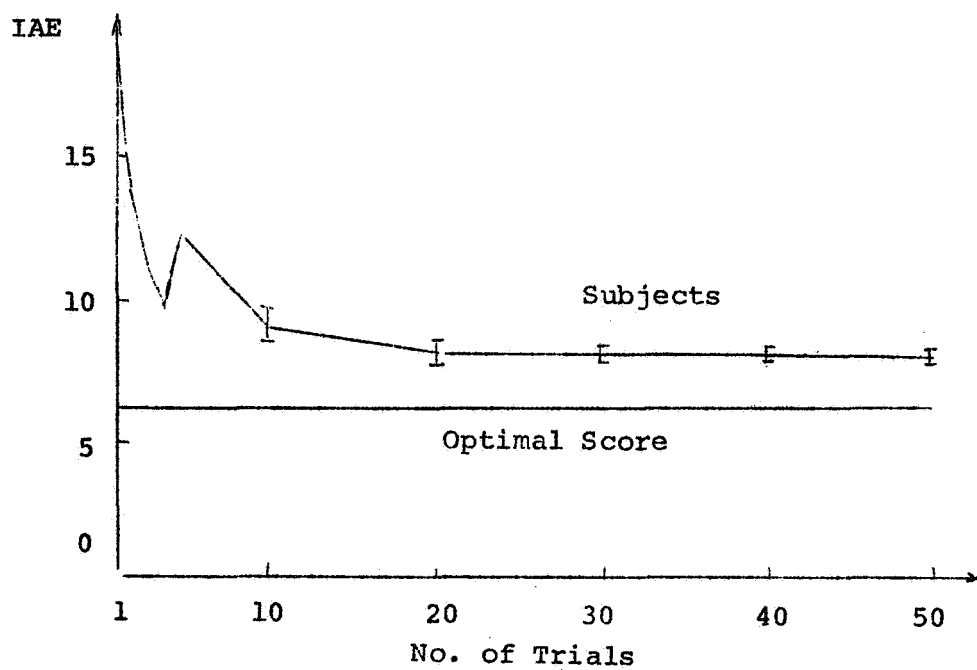


Figure 2. —Unstable plant learning curves.

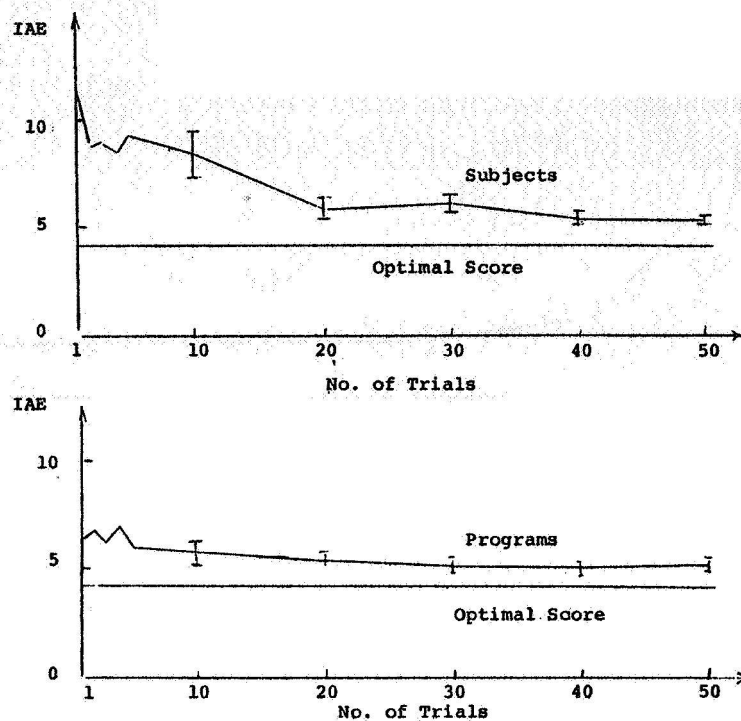


Figure 3.—Harmonic oscillator learning curves ($\omega = 1$ rad/sec).

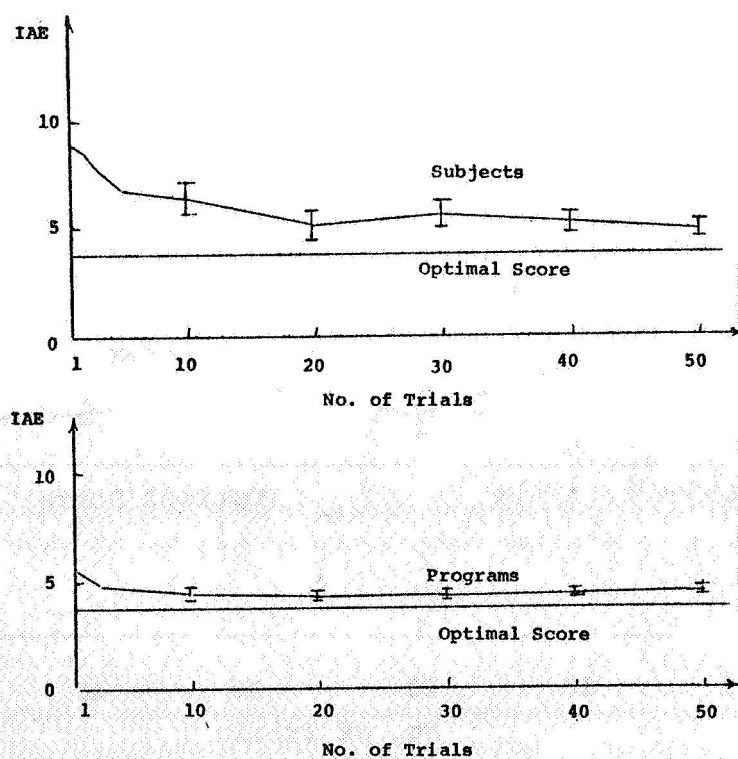


Figure 4.—Harmonic oscillator learning curves ($\omega = 1$ rad/sec).

REFERENCES

1. Preyss, A. E.; and Meiry, J. L.: Stochastic Modeling of Human Learning Behavior. Third Annual NASA-University Conference on Manual Control. SP-144, 1967.
2. McRuer, D. T.; Graham, D.; Krendel, E.; and Reisener, W. C.: Human Pilot Dynamics in Compensatory Systems—Theory, Models, and Experiments with Controlled Element and Forcing Function Variations. AFFDL-TR-65-15, July 1965.
3. Athans, M.; and Falb, P. L.: Optimal Control—An Introduction to the Theory and Its Applications. McGraw-Hill Book Co., Inc., 1966.
4. Mann, H. B.; and Whitney, D. R.: On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. Ann. Math. Stat., vol. 18, 1947, pp. 50-60.
5. Chien, T. T.: Human Learning Behavior in Manual Control Tasks. S.M. Thesis, MIT, Sept. 1967.

18. Contributions of Human Behavior as a Multiloop Controller

*G. Schweizer
Friedrichshafen, Germany*

Experiments were conducted to review the existing describing function models on human behavior for single channel tasks. Random appearing stationary forcing functions and disturbances were present and it was assumed that the operator established stationarity during all the tests. The results of the analyzed data indicated that a linear model with random time-varying coefficients provided a good description of the operator's behavior.

Studies were conducted to determine what modifications to the single variable model system are necessary for a good representation of multivariable manual control systems. The experiments show that a sample data system with random sample intervals seems to be necessary.

This model becomes complicated. Therefore, the application of the concept of information theory has been tried to assess the performance of the human operator in multiloop tasks. It was learned that with the knowledge of the internal dynamics of the operator, there is some hope of determining performance indexes for manual control systems with the information theory method.

INTRODUCTION

The investigations reported here are an element of a research program to explore adequate means to describe the human behavior as a controller. The primary objective of the studies so far has been to achieve reasonable descriptions for man-machine systems with visually presented random-appearing stationary forcing functions and disturbances. For the present the central aim is to explore the human operator's behavior in special tasks where an attempt is made to keep all the environmental factors constant during one run. Furthermore it is assumed that the operator himself established stationarity during an experiment.

The investigations required a refinement of the quasi-linear model, aimed to include the linear time-varying behavior of the human operator. As the feedback-control theory models for multiloop tasks become quite complicated, an approach has been tried to use the elementary information theory for the description.

REVIEW OF THE FEEDBACK-CONTROL MODELS FOR SINGLE LOOP MAN-MACHINE SYSTEMS

Virtually all of the research on the human operator has dealt with single loop control systems. The model of figure 1 has been investigated by numerous investigators. This model was refined during the present program because it provided no explanation for a number of physical observations. The major correction of the existing model was the introduction of random time-varying coefficients.

The corrected model shown in figure 2 avoids some of the observed data deficiencies. The operator's behavior dealing with the control of stationary forcing functions and disturbances in a one-dimensional control task is described by the differential equations

$$\dot{y}(t) = e^{-\gamma(t)} [t - T_T - \gamma(t)] \quad (1)$$

$$\sum_{i=0}^N a_i k^{(i)}(t) = \sum_{i=0}^M b_i \dot{y}^{(i)}(t) \quad (2)$$

$$-\sum_{i=0}^M \beta_i(t) \dot{y}^{(i)}(t) = \sum_{i=0}^N \alpha_i(t) k^{(i)}(t)$$

$e(t)$ is the visual presented random input information, $k(t)$ the total operator output, a_i and b_i are constant coefficients. $\alpha_i(t)$, $\beta_i(t)$, and $\gamma(t)$ are stationary normal distributed random functions with $\gamma(t) < T_T$.

In a single loop control loop in figure 3 the refined model comprises two relations of practical interest, if the system is stable

$$\frac{S_{ek}(\omega)}{S_{ee}(\omega)} = e^{-j\omega T_T} H_1(j\omega) \quad (3)$$

$$S_{kk}(\omega) = |H_1(j\omega)|^2 \cdot S_{ee}(\omega) + S_{pp}(\omega) \quad (4)$$

with the notion

$$H_1(j\omega) = \frac{\sum_{i=0}^M b_i (j\omega)^i}{\sum_{i=0}^N a_i (j\omega)^i}$$

$S_{ee}(\omega)$ is the power spectrum of the input,

S_{ek} the cross power spectrum between the input and the output signal.

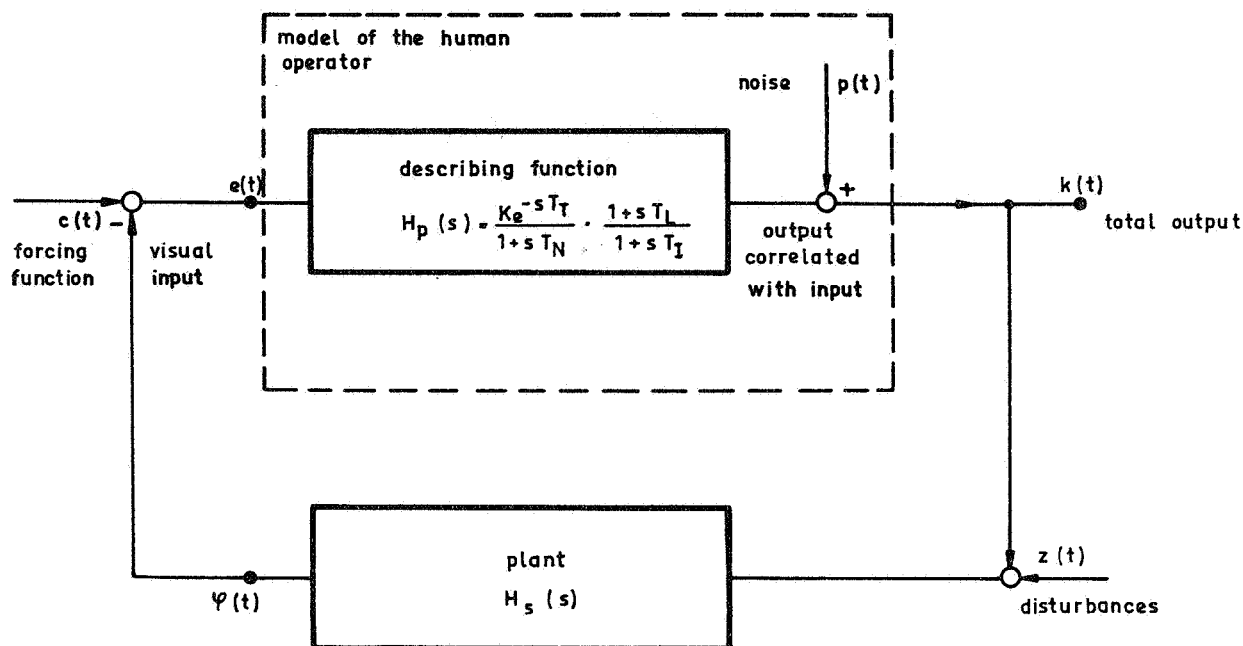


Figure 1.—Man-machine feedback system.

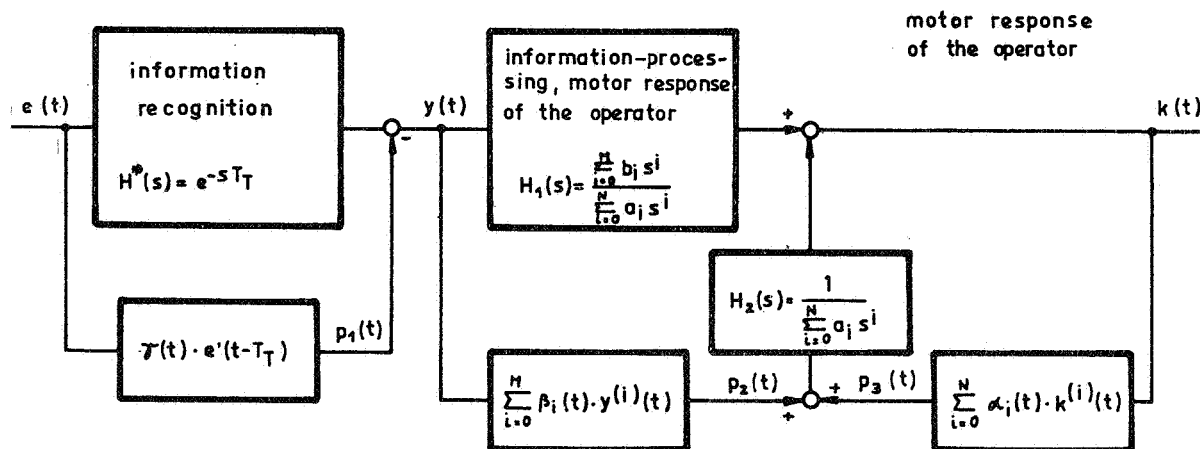


Figure 2.—Model of the human operator with random varying coefficients.

$S_{pp}(\omega)$ is the power spectrum related to the random time varying coefficients. Equations (3) and (4) are of practical interest in the areas of the analysis of in-flight data, hence no external artificial signal is required for many man-machine investigations. It is emphasized that there are basic differences in the models of figures 1 and 2 in spite of some similar elements. The model in figure 2 explains the remnant as parametric excitations caused by random time-varying behavior of the operator, whereas the remnant in figure 1 is assumed to be superimposed on the operator's output. It would be beyond the framework of this report to explain all the different conclusions drawn from the two models. But it is noted here that the critical factor in determining the man-machine system instability with the model of figure 3 is not only the describing function $H_1(j\omega)$ as in figure 1 but is the magnitude of the time-varying coefficients as well.

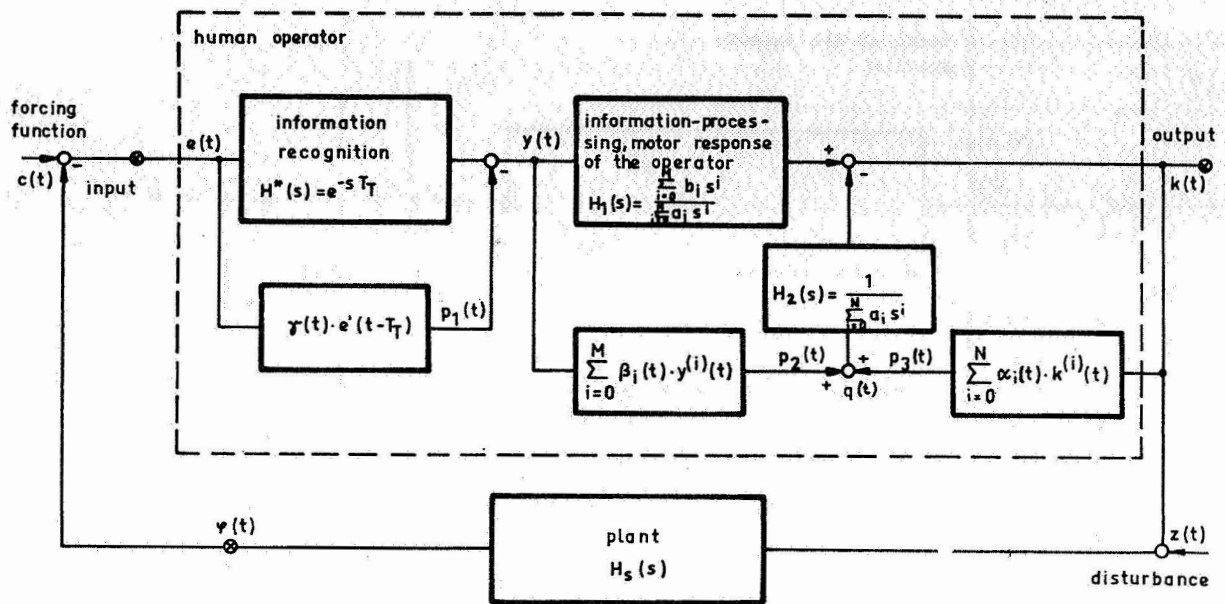


Figure 3.—Feedback model of a man-machine system with varying random coefficients.

Hence results based on a crossover model to investigate man-machine system instabilities can become doubtful, if the model of figure 2 proves to be an adequate description for the operator's performance. The model of figure 2 has been further refined to investigate whether nonlinear effects were playing a major role during the experiments. Equation (2) has been modified to include nonlinear terms.

$$\begin{aligned}
 \sum_{i=0}^N a_i k^{(i)}(t) = & \sum_{i=0}^M b_i y^{(i)}(t) - \sum_{i=0}^M \beta_i(t) y^{(i)}(t) - \sum_{i=0}^N \alpha_i(t) k^{(i)}(t) \\
 & - \underbrace{\sum_{i=0}^M \varepsilon_i \left[y(t), y'(t), \dots, k(t), k'(t), \dots \right] y^{(i)}(t)}_{m(t)} \\
 & - \underbrace{\sum_{i=0}^N r_i \left[k(t), k'(t), \dots, y(t), y'(t), \dots \right] k^{(i)}(t)}_{n(t)} \quad (5)
 \end{aligned}$$

The nonlinear terms in equation (5) can be approximated by power series in a final interval

$$\begin{aligned}
 f_i = & \sum_{\mu=0}^N a_{i\mu} k^{(\mu)}(t) + \sum_{\mu=0}^M b_{i\mu} y^{(\mu)}(t) + \\
 & + \sum_{\mu=0}^N \sum_{v=0}^N a_{i\mu v} k^{(\mu)}(t) \cdot k^{(v)}(t) + \sum_{\mu=0}^N \sum_{v=0}^M a_{i\mu v} k^{(\mu)}(t) \cdot y^{(v)}(t) + \\
 & + \sum_{\mu=0}^M \sum_{v=0}^M b_{i\mu v} y^{(\mu)}(t) \cdot y^{(v)}(t) + \\
 & + \sum_{\mu=0}^N \sum_{v=0}^N \sum_{\rho=0}^N a_{\rho\mu v} k^{(\mu)}(t) \cdot k^{(v)}(t) \cdot k^{(\rho)}(t) \cdot \dots
 \end{aligned} \tag{6}$$

It can be shown by theoretical investigations of equations (5) and (6) that the operator's behavior is linear during a run with random stationary normal distributed forcing functions and disturbances, if

$$\begin{aligned}
 E \{k(t+\tau) \cdot e(t)\} &= \phi_{ke}(\tau) \\
 E \{k(t+\tau) \cdot e^2(t)\} &= 0 \\
 E \{k(t+\tau) \cdot e^3(t)\} &= 3\phi_{ee}(0) \phi_{ek}(\tau) \\
 E \{k(t+\tau) \cdot e^4(t)\} &= 0
 \end{aligned} \tag{7}$$

Experimental results have shown that the operator performed mainly as a linear controller, if he had to deal with random-appearing stationary signals. During the runs all environmental factors were kept constant:

FEEDBACK MODELS OF MAN-MACHINE SYSTEMS FOR MULTILoop TASKS

A simulator and a rig program with a flying bedstead shown in figure 4 were aimed to get insight into the human operator's behavior as a multiloop controller. Figures 5 and 6 show the simplified block diagrams. In the case of the simulated helicopter we are dealing with an intercoupled multivariable plant, whereas the flying rig represents to a first approximation four decoupled control loops, if the translationary motions are neglected.

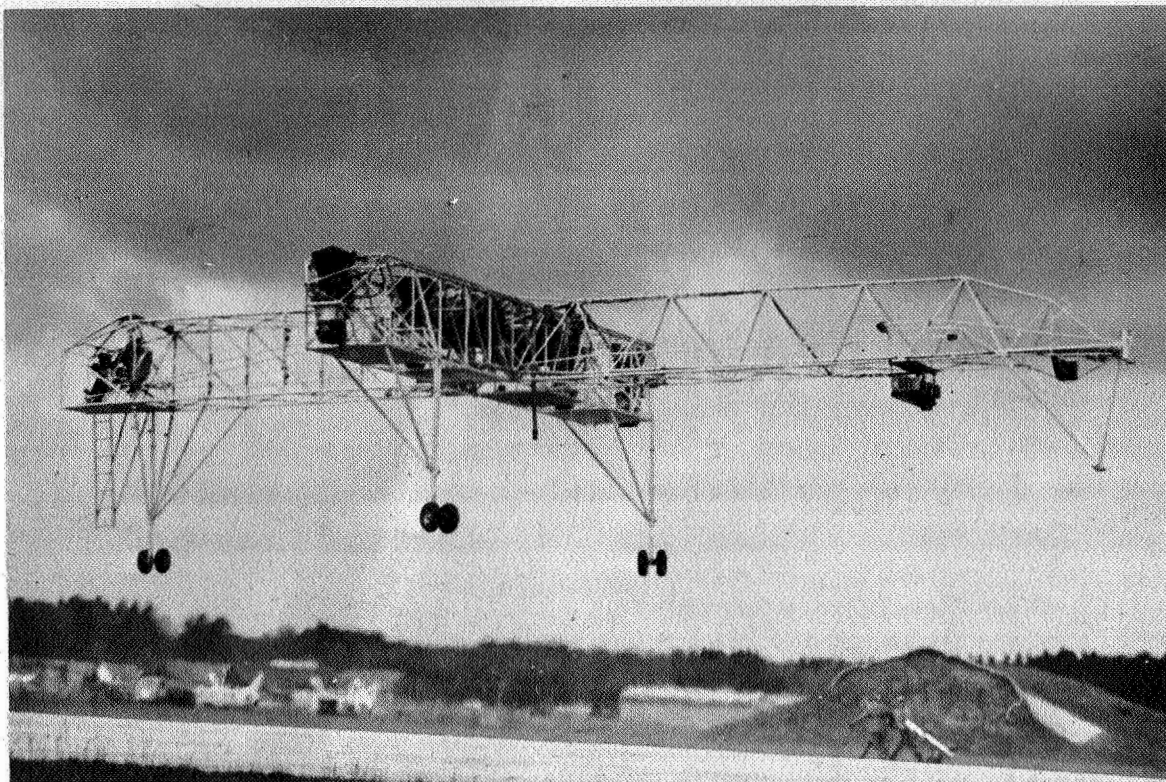


Figure 4. — Flying bedstead.

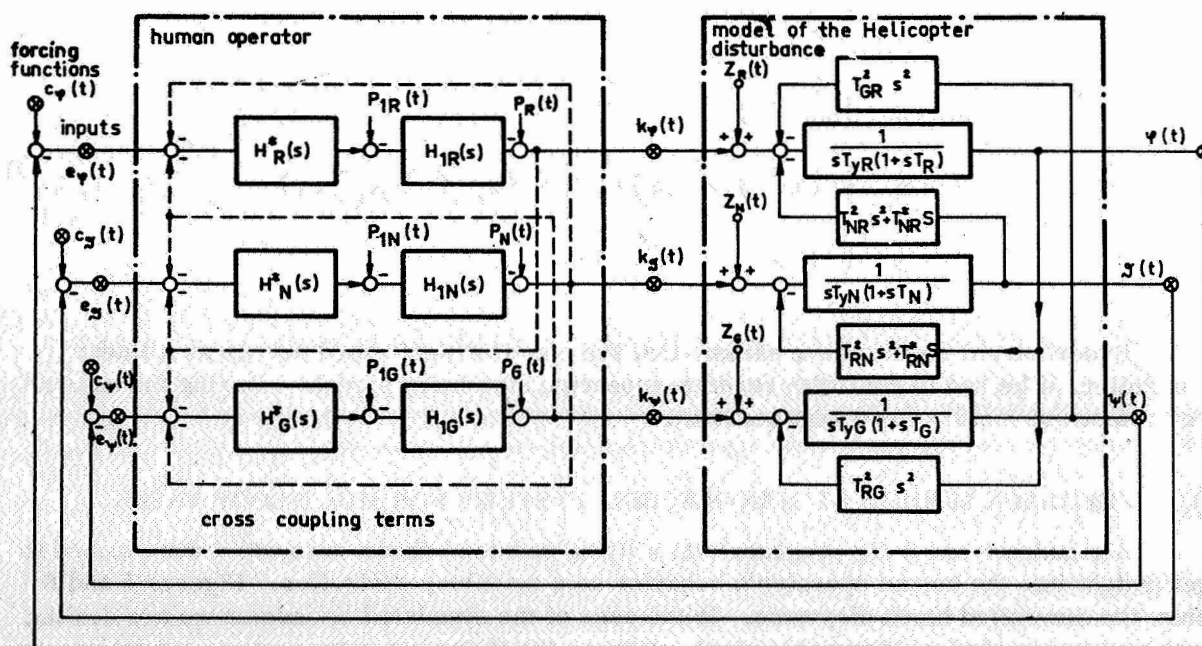


Figure 5. — Block diagram of the simulated manual controlled helicopter.

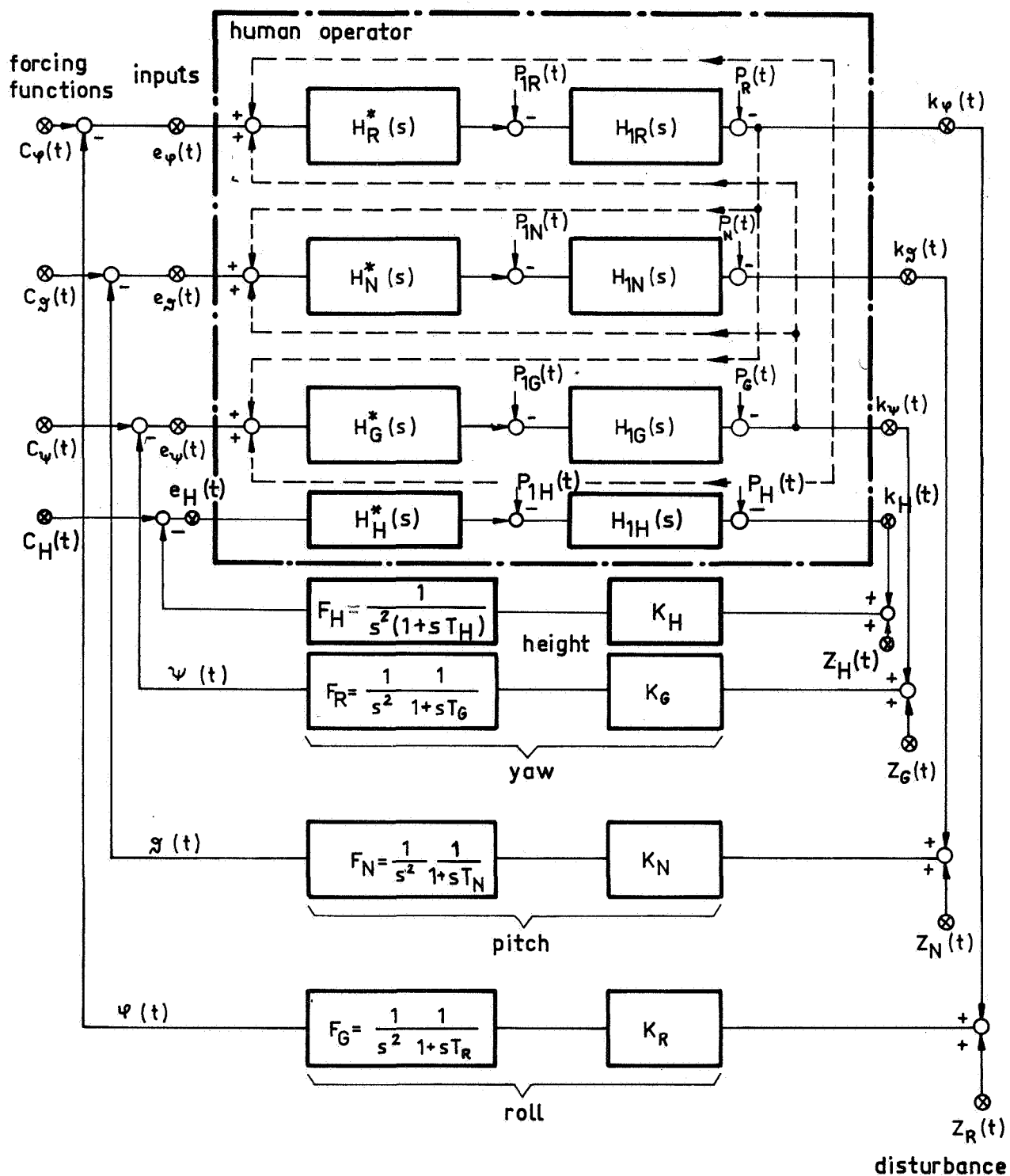


Figure 6. —Block diagram of the manual controlled rig.

TABLE 1.—CHANGES IN HUMAN OPERATOR'S
DESCRIBING FUNCTION

Vehicle dynamics	C.S. output ^a	Percent improvement in rise ^b	Phase change at 4.3 rad/sec ^c	Phase change at 7.7 rad/sec ^d
$\frac{e^{-0.1s}}{s}$	10	38	8	40
$\frac{20e^{-0.1s}}{s(s+20)}$	10	35	11	30
$\frac{4e^{-0.1s}}{s(s+4)}$	10	29	8	38
$\frac{2e^{-0.1s}}{s(s+2)}$	10	30	13	40
$\frac{e^{-0.1s}}{s(s+1)}$	10	20	0	40
$\frac{e^{-0.1s}}{s^2}$	10	19	13	38
$\frac{4e^{-0.1s}}{s^2(s+4)}$	20	0	14	30
$\frac{2e^{-0.1s}}{s^2(s+2)}$	20	0	16	20
$\frac{5e^{-0.1s}}{s^2+3s+5}$	10	50	7	50
$\frac{5e^{-0.1s}}{s(s^2+3s+5)}$	5	0	14	42

^aThe maximum control stick output voltage.

^bThe improvement in the system relative integral squared error when motion cues are added to visual cues.

^cThe increase in the phase lead of the human operator's describing function at 4.3 rad/sec when motion cues are added to the visual cues.

^dThe increase in the phase lead of the human operator's describing function at 7.7 rad/sec when motion cues are added to the visual cues.

During the rig tests the pilot had mainly to control four variables: roll, pitch, yaw, and height by stick, pedals, and throttle movements. All input information was displayed visually. The analysis of the data obtained during many runs indicated that the time-varying model of figure 2 could be used to explain many observations for the individual channels. Input and output power spectra and the describing functions for the different loops have been measured. In addition cross-correlation measurements have been performed between the outputs of the different channels of the multiloop human operator to get information as to whether any cross-coupling terms have been present. It was learned that no significant coupling could be detected between the different channels of the multiloop model of the operator. Figure 7 shows a representative result.

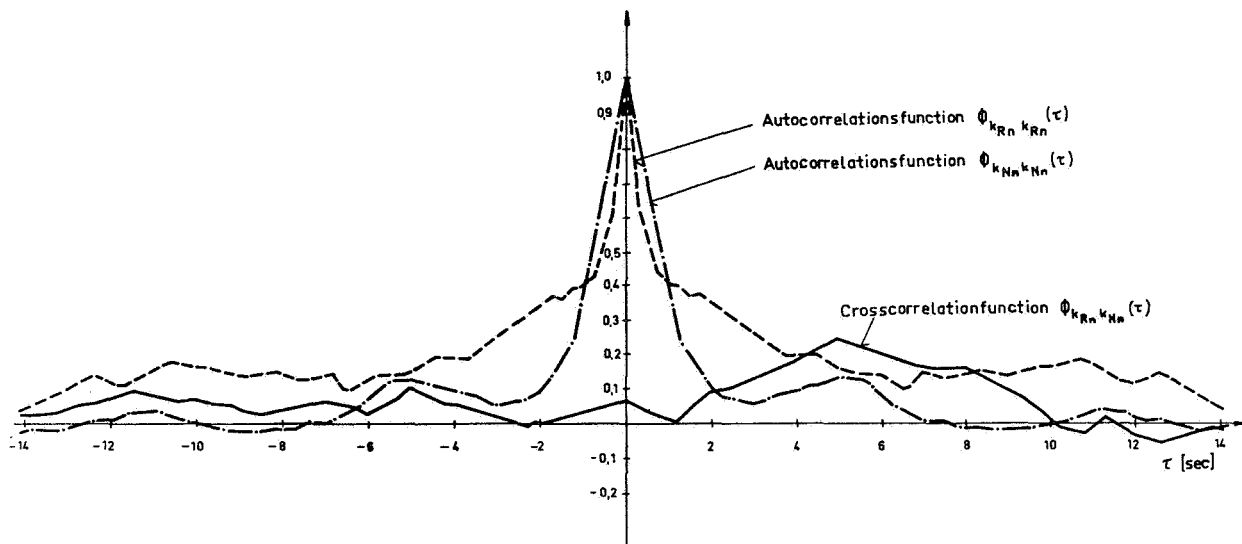


Figure 7. —Normalized auto and cross-correlation function of the motor response for roll and pitch control of the rig.

During the simulated helicopter tests the pilot had to control the roll, pitch, and yaw angle by stick and pedal movements. The necessary information was displayed visually with the aid of a pseudoperspective pattern. The feedback model of figure 2 could be used to explore details of the human operator's behavior controlling the three different channels. Figure 8 shows some describing functions and some power spectra, which have been obtained during different runs. The data are representative of the majority of the numerous results from the whole program. If the model of figure 5, assuming a continuous dynamic model for the operator and an intercoupled plant, is valid and there are no coupling terms within the model of the human operator, the cross correlation must be nonzero because of the coupling terms of the controlled element. The same is true if there are coupling terms within the model for the operator, assuming that these are not chosen to achieve non-interaction. The analysis of the data indicates that the correlation between the different output signals varies in an arbitrary manner as shown in figure 9. This observation can be explained by choosing a sample data model for the human operator. However, such a model becomes quite complicated without providing much insight.

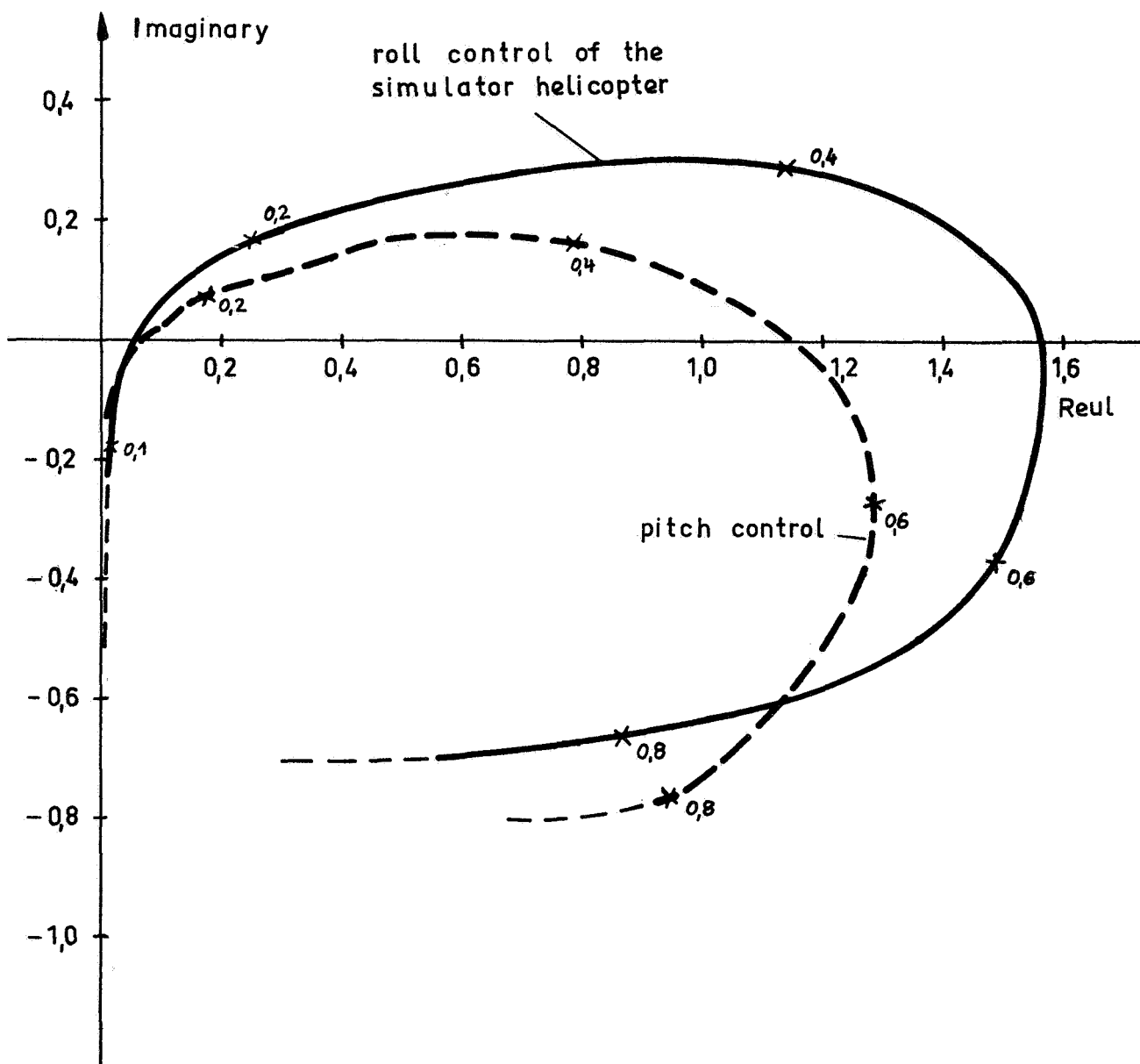


Figure 8(a).—Describing functions of the human operator.

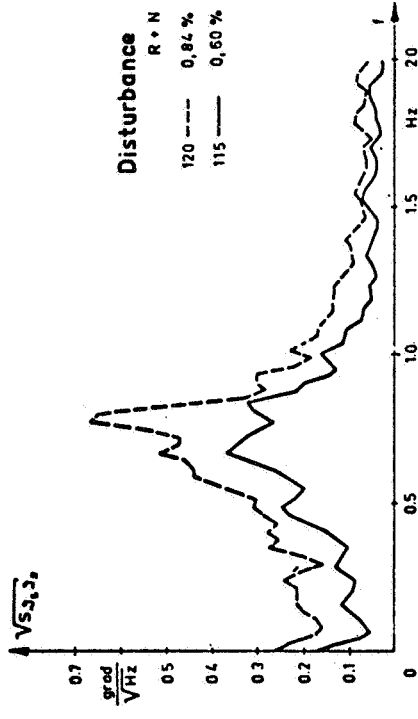


Figure 8(b). —Power spectra for roll control.

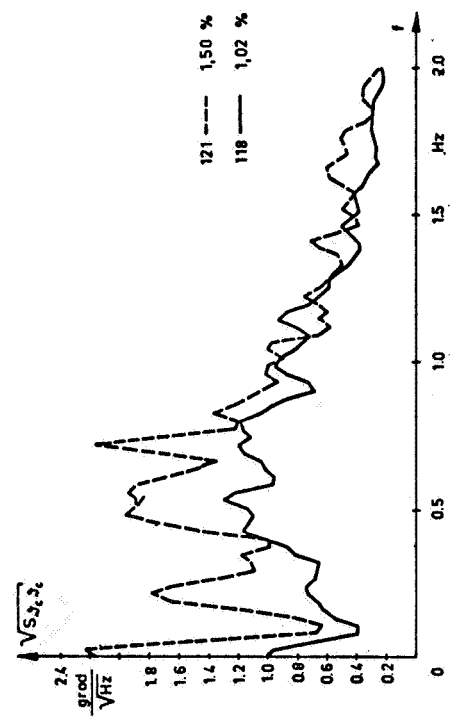


Figure 8(c). —Power spectra for pitch control.

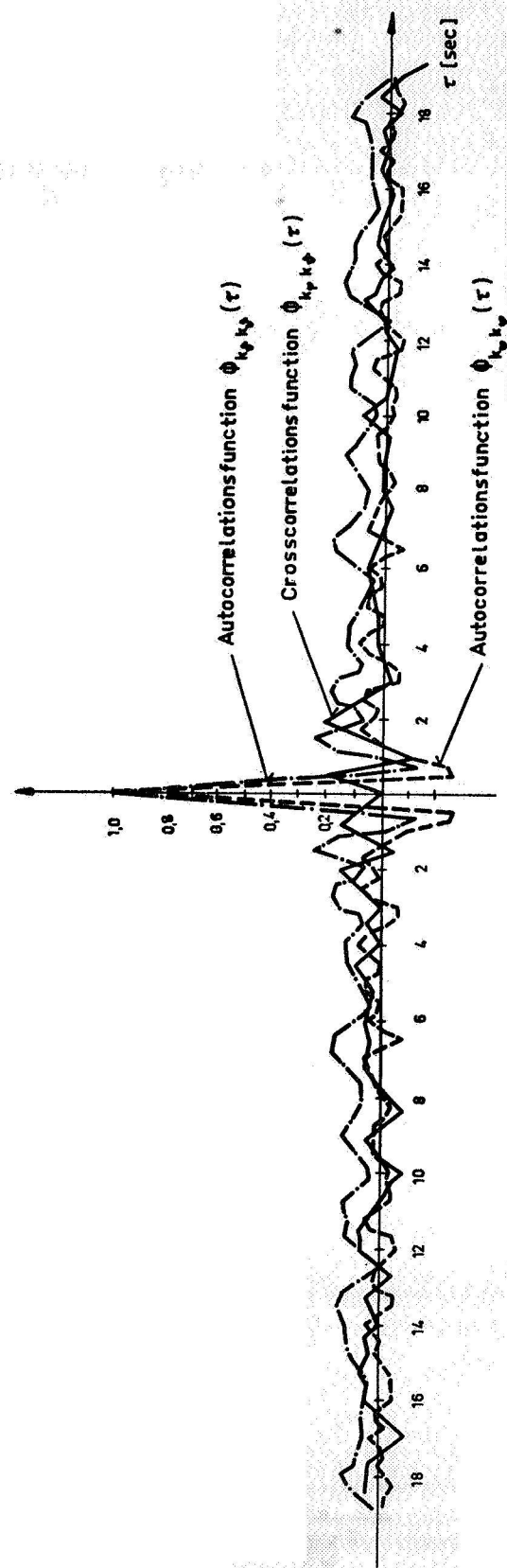


Figure 9(a). — Normalized auto and cross-correlation function of the motor response for pitch and roll control of the simulated helicopter.

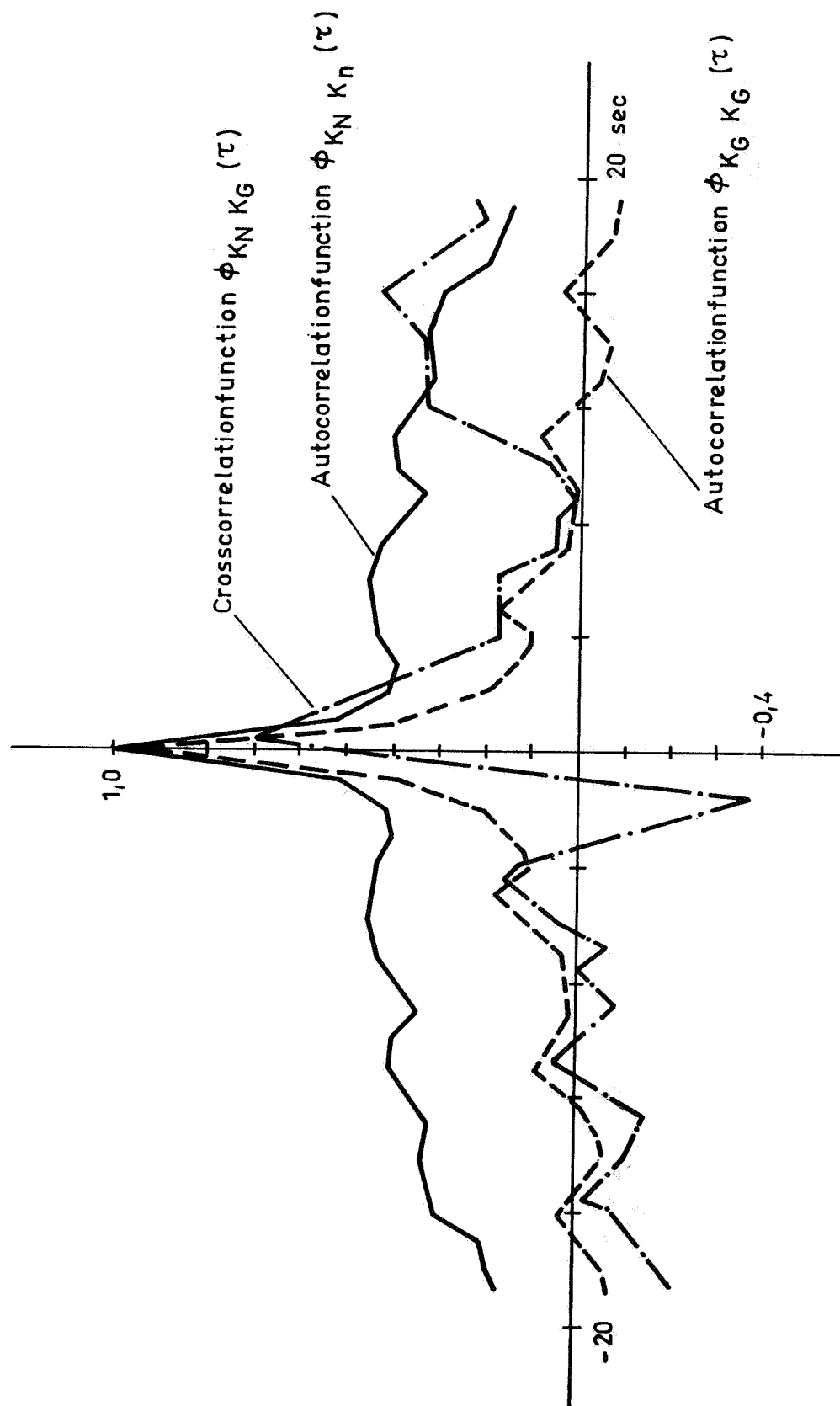


Figure 9(b). — Normalized auto and cross-correlation function of the motor response for yaw and pitch control of the simulated helicopter.

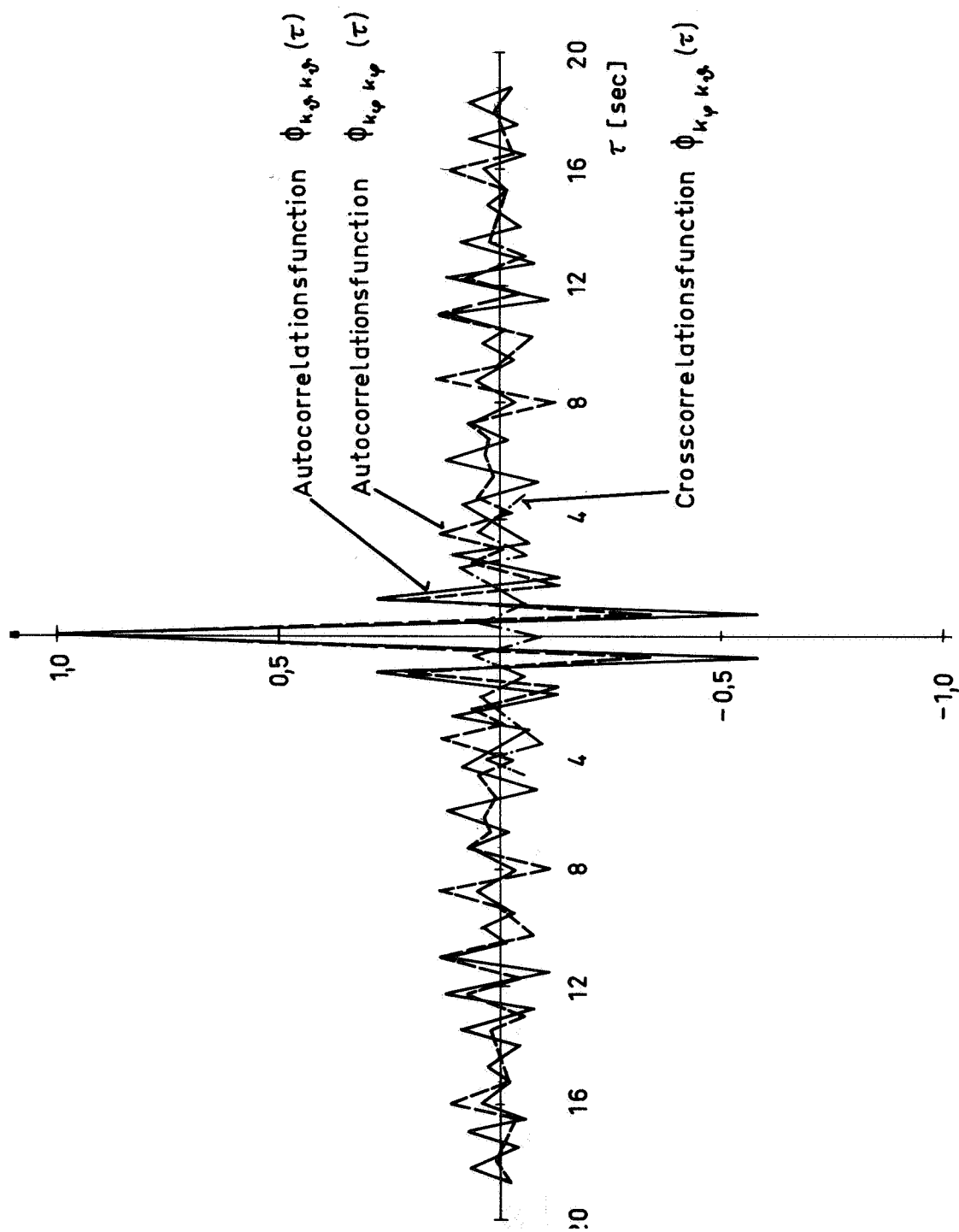


Figure 9(c). —Normalized auto and cross-correlation function of the motor response for pitch and roll control of the simulated helicopter.

THE USE OF ELEMENTARY INFORMATION THEORY FOR THE DESCRIPTION OF THE HUMAN OPERATOR IN MULTILoop TASKS

The output $k(t)$ of the model of the human operator contains processed information from the input and signals $p(t)$ not correlated with the input. We define the part of the output correlated with the input signal as processed information and $p(t)$ as permissible error signals. The information generation rate of the human operator as a single channel controller can thus be described by

$$IF = f_T \log_2 \frac{\overline{k^2}}{p^2} \left[\frac{\text{bit}}{\text{sec}} \right] \quad (8)$$

f_T is the cutoff frequency of the signal $k(t)$.

As a multiloop controller the information generation rate is

$$IF = \sum_{i=0}^m T_{Tn} \log_2 \frac{\overline{km^2}}{p_m^2} \left[\frac{\text{bit}}{\text{sec}} \right] \quad (9)$$

$$T/T_0 = Cf_T \log \frac{\overline{k^2}}{p^2} \quad (10)$$

For multivariable tasks we obtain

$$T/T_0 = \sum_{i=0}^n c_i f_{T_i} \log_2 \frac{\overline{k_i^2}}{p_i^2} \quad (11)$$

The information generation rate can be measured for each channel if the operator has to deal with a multivariable control task. Table 1 shows results from rig and simulator tests.

It was learned during the tests that the information generation rate required for control depends mainly on the human operator's environment and on the parameters of the controlled elements. The same is true for the maximal information rate which the operator can process during a given task. The performance of the operator was marginal during run no. 2, where the pilot had to control a single channel, whereas he was not yet at his limits during run no. 5. This was a free flight with the flying bedstead where the pilot had to control six degrees of freedom. The total information flow processed by the human operator during these multivariable control tasks was considerably higher, compared with the single channel experiment no. 2. During the simulator tests with the helicopter the pilot reached his limits when he had to control three degrees of freedom as shown in the block diagram.

The results indicate clearly that the ability to process information during a control task depends very much on environmental factors and on the element to be controlled. However,

TABLE 1.—RESULTS OF RIG AND SIMULATOR TESTS

run	rig-test on the pedestrial	controlled element	control-loop	information generation rate	remark
1		$\frac{K}{S^2}$ (K adjusted to the optimal operator performance)	roll axis (single channel task)	~2.7 bits/sec	control was satisfactory but marginal
2	with artificial introduced disturbances single channel (yaw and pitch loop fixed)	$\frac{K}{S^2}$ (K not adjusted to the optimal opera- tors performance)	roll axis (single channel task)	~3.0 bits/sec	control was unsatisfactory and marginal
3		$\frac{K}{S(ST + 1)}$ (K optimal adjusted)	roll axis (single channel task)	~2.0 bits/sec	control was satisfactory
4		$\frac{K}{S(ST + 1)}$ (K not optimal adjusted)	roll axis (single channel task)	~2.6 bits/sec	control was satisfactory but marginal

TABLE 1. --CONCLUDED

5	six degrees of freedom tasks partially evaluated	$\frac{K_R}{s^2}$ (K adjusted)	roll axis	2.2 bits/sec	six degrees of freedom control was satisfactory but marginal
		$\frac{K_N}{s^2}$ (K adjusted)	pitch axis	2.2 bits/sec	
6	(see block diagram 6)	$\frac{K_R}{(s^2 T_{1R}^2 + s T_{2R} + 1)}$	roll axis	0.8 bits/sec	
		$\frac{K_N}{(s^2 T_{1N}^2 + s T_{2N} + 1)}$	pitch axis	0.7 bits/sec	
7	simulated helicopter tests with artificial introduced disturbances	intercoupled plant see fig 5	roll axis	2.0 bits/sec	three degrees of freedom control was marginal
			pitch axis	1.8 bits/sec	
			yaw axis	1.8 bits/sec	

whether a single channel or a multivariable control task has to be performed seems of no effect on the human operator's capability.

A factor C with the dimension time per bit is introduced to describe the operator's performance for the control of a variable in a single loop or in a multivariable loop. With this we obtain a relation which gives the ratio of a timespan required for the necessary information generation to control a variable to the time unit T_0 which is available.

The factor C was to be in the order of 0.3 bit/sec during the runs 1 to 4 and approximately 0.15 to 0.2 bit/sec during the runs 5 to 7.

There are a variety of different means to ease the human operator's job. With automatic control equipment usually f_{Ti} becomes smaller and by properly displaying the input information one achieves lower values of C_i .

CONCLUSIONS

Feedback models for the description of multivariable control tasks of a human operator become complicated. The elementary information theory provides an additional adequate mean to characterize the potential performance of the human operator. However, stability of the man-machine system is a prerequisite for the use of the methods of information theory as described above. To evaluate equations (8) to (11), f_{Ti} , k^2 , and p_1^2 have to be computed. For this at least a model for a single channel is required. The model of figure 2 has proved to be most useful to process in-flight data.

19. Pilot's Response to Stability Augmentation System Failures and Implications for Design*

*David H. Weir and Walter A. Johnson
Systems Technology, Inc.*

The dynamic response of the human pilot is studied during sudden changes in the effective controlled dynamics caused by stability augmentation system failure. Experimental results from multiple-loop fixed-base studies are presented. A hypothesis of graceful degradation is shown to be valid which states that the pilot's transition response and performance are improved if the difference in controlled element dynamics at failure is reduced. The design implications of this principle are detailed. A model for the pilot's dynamic response is presented which accounts for his behavior during the several phases of transition.

INTRODUCTION

Failure of a vehicle's stability augmentation system can cause a sudden change in the effective controlled element dynamics and require the human pilot to retain control while adapting to the new dynamics or taking corrective measures. Although the problem can be avoided by use of fail-operational systems, an attractive alternative in some cases is to use the pilot's control capability plus reduced system redundancy to achieve a design compromise with its attendant savings in weight, cost, etc. To use such a mix of manual and automatic functions, the pilot's capability must be understood and the effective vehicle dynamics must be tailored to make the system degradation at failure as graceful as possible.

Past studies have considered controlled element transitions in single-loop tracking tasks (refs. 1 to 6) and have concentrated on failure identification and the pilot's dynamic response. The current results emphasize multiple-loop lateral/directional control of a conventional airframe with a yaw damper failure which results in an abrupt or steplike change in the dynamics. Of central concern is the pilot's ability to retain control at failure and the degree to which he can successfully and reliably do so can influence the design philosophy of the augmentation system. If a basically unstable vehicle is augmented to give very good handling, then the transition due to failure will require a large increment in his readaptation and pilot/vehicle performance may be very poor. If, instead, the prefailure augmentation is reduced to place modest demands on the pilot at all times, the increment of readaptation will be smaller and the pilot should perform more capably if and when a failure

*This research was accomplished in part under NASA Contract NAS2-3607.

occurs. These notions lead to the graceful degradation hypothesis which states: The pilot's transition response and performance are a function of the difference in controlled element dynamics across the transition as well as their absolute forms; the larger the difference the greater the control difficulty and the poorer the performance.

EXPERIMENTS

The basic experimental task was roll control in a fixed-base simulator with a center stick and a simulated gust input. The failure task involved pilot control with rudders of the "Dutch roll" mode. The roll dynamics were contrived to give good handling in all configurations. The Dutch roll dynamics were varied from good to poor by changing the level of prefailure augmentation. No manual rudder was required in the good configuration, whereas various amounts were needed in the poorer ones. At failure, the effective airframe dynamics went from good to bad or poor to bad. Both hardover and soft failure modes were simulated. The principal performance measures were deterministic descriptors of the system error following failure; e.g., divergence rate, peak error, and settling time.

The basic airframe was one that the pilot had to fly with manual rudder plus aileron in the event of an augments failure. The following lateral/directional characteristics were sought:

Neutrally stable spiral mode

Good roll subsidence

Adverse ω_ϕ/ω_d to force pilot rudder control

Negative Dutch roll damping to make the task more demanding

Dutch roll frequency within the pilot's bandwidth; e.g., about 2 rad/sec

Good $r \rightarrow \delta_r$ dynamics

Reasonably high $|\phi/\beta|$ at Dutch roll frequency

The synthesized airframe had the following transfer functions:

$$\frac{\phi}{\delta_a} = \frac{-40[s^2 + 2(-.13)(2.29)s + (2.29)^2]}{(s - .024)(s + 5)[s^2 + 2(-.14)(2)s + (2)^2]}$$

$$\frac{r}{\delta_r} = \frac{-3.2(s + 5)[s^2 + 2(-.04)(.2)s + (.4)^2]}{(s - .024)(s + 5)[s^2 + 2(-.14)(2)s + (2)^2]}$$

The gains were selected for good subjective opinion.

To test the graceful degradation hypothesis, three prefailure, augmented airframes were used; configuration A with good to optimum Dutch roll damping, configuration B with only slight Dutch roll damping and a value of ω_ϕ/ω_d equal to unity, and configuration C with slight damping and ω_ϕ/ω_d equal to 1.3. Closure of the augments loop alone did not provide all of the desired configurations; therefore, the stability derivatives N'_r and N'_{δ_a} were changed as the augments loop was closed. They were changed back instantaneously when a configuration was failed to the basic airframe, allowing ω_ϕ/ω_d and ζ_ϕ and ζ_d to be varied independently of any other parameters. To keep the augments from opposing the pilot, a washout was included with a time constant of 0.25 second. The resultant bank angle control transfer functions for the augmented airframe were—

Configuration A (good):

$$\left(\frac{\phi}{\delta_a}\right)_{r \rightarrow \delta_{rSAS}} \doteq \frac{-40(s + .31)[s^2 + 2(.70)(1.80)s + (1.80)^2]}{s(s + .52)(s + 5)[s^2 + 2(.70)(1.65)s + (1.65)^2]}$$

Configuration B (fair, $\omega_\phi \doteq \omega_d$):

$$\left(\frac{\phi}{\delta_a}\right)_{r \rightarrow \delta_{rSAS}} \doteq \frac{-40(s + .26)[s^2 + 2(.16)(1.96)s + (1.96)^2]}{s(s + .29)(s + 5)[s^2 + 2(.15)(1.95)s + (1.95)^2]}$$

Configuration C (poor, $\omega_\phi > \omega_d$):

$$\left(\frac{\phi}{\delta_a}\right)_{r \rightarrow \delta_{rSAS}} \doteq \frac{-40(s + .26)[s^2 + 2(.13)(2.52)s + (2.52)^2]}{s(s + .29)(s + 5)[s^2 + 2(.15)(1.95)s + (1.95)^2]}$$

Configuration A was contrived so that the pilot could use aileron-only control before the failure, but had to use the rudder after the failure. In B the pilot needed the rudder occasionally before the failure if the Dutch roll was excited by the disturbance. With C some manual rudder control was always required before failure to avoid destabilizing the Dutch roll. The system block diagram is shown in figure 1. Representative values of the valve-actuator and rate gyro dynamics were used.

Two kinds of rudder transients were used to simulate hard and soft failure modes, and these were applied at the instant of augments failure. The hard-failure transient consisted of a washed-out rudder step, while the soft-failure was simulated by a "hold" of the rudder signal from the augments. As a result of the hold there was no disturbance imposed on the airplane by the failure, and the pilot had no immediate cue in the soft case to indicate that the augments had failed. The two kinds of failures were presented with equal frequency in a random sequence.

The cockpit was fixed base with a conventional seat, stick, and pedal arrangement. An opaque canopy was closed during the experiments to isolate the pilot from the laboratory. The primary displays of bank angle and turn rate were generated on a two-gun CRT located roughly at eye height and about 12 inches from the pilot. An additional meter was set up to indicate lagged sideslip. The displays were arranged as shown in figure 2.

The airplane was disturbed by a random-appearing sideslip gust input consisting of the sum of sine waves shown in table 1. The effective displayed bandwidth differed from that shown in table 1 because the gust was filtered by the airframe dynamics. The rms gust amplitude was 1.4° sideslip.

Two subjects were used in the experiments, one a high time military pilot and the other a private pilot. They were instructed to:

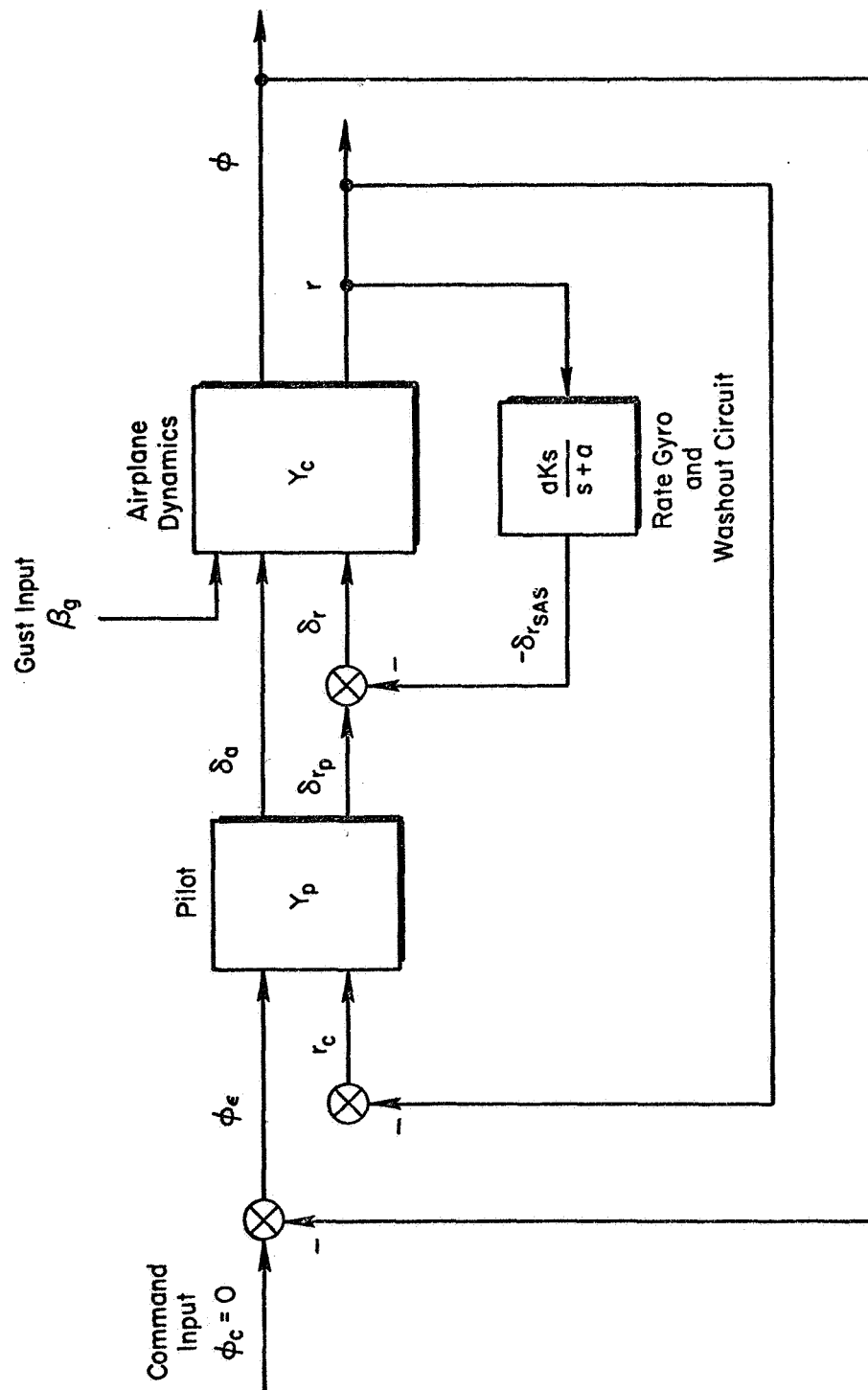


Figure 1. — Block diagram of multiple-loop control task.

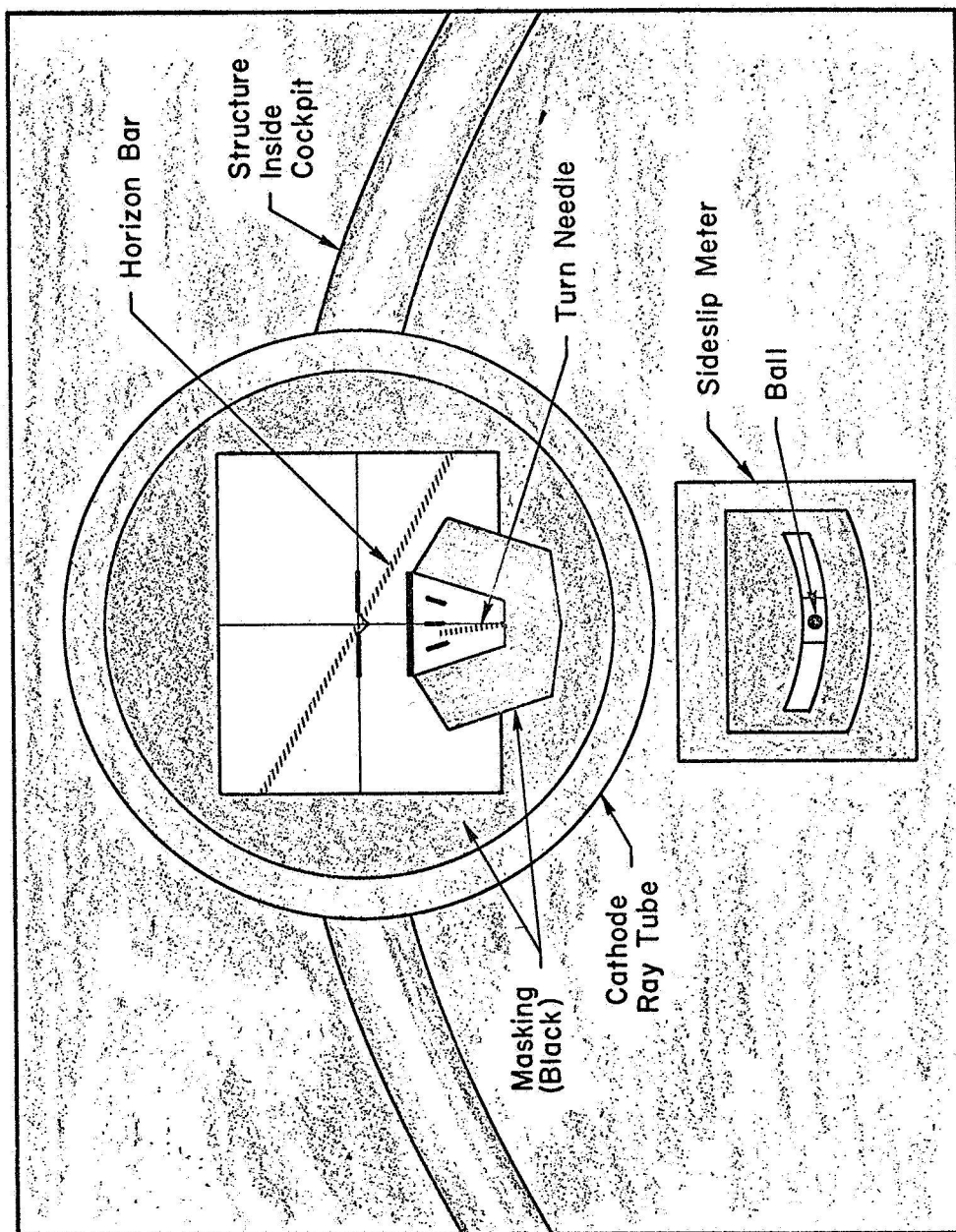


Figure 2. — Cockpit display (indicating coordinated left turn).

Minimize bank angle and yaw rate as well as you can throughout the run. Try to never lose control of the vehicle after a failure. If at any time you feel tired, we would like you to stop and take a rest so we can get consistent data from an alert pilot.

The pilots evolved their own error minimization strategy. They practiced the failure tasks 2 hours a day for several days, until each had accomplished about 200 trials over all configurations. Subsequent to the learning period, their performance and response was substantially the same in a given task. The experimental runs lasted about 2 minutes each, with the augments failure occurring after about 1 minute. The runs were grouped in consecutive sets of five (with the same prefailure dynamics), followed by a rest period between sets. The data of interest here consist of 90 experimental runs for the military pilot, recorded on magnetic tape and strip charts. This provides 15 replications for both the hard and soft modes with each of the three prefailure configurations.

DATA ANALYSIS

Prior research (ref. 1) has shown that the operator's response to a controlled element transition contains four general phases:

- Prefailure steady state; stationary compensatory tracking
- Retention of prefailure conditions
- Nonlinear control; large control actions that stabilize the system and reduce the error to some acceptable level
- Postfailure steady state; stationary compensatory tracking

The retention phase starts at failure and consists of the prefailure pilot adaptation and the postfailure controlled element operating in a closed-loop fashion on the system error. This frequently results in an unstable condition because the typical failure results in a

TABLE 1.—SUM OF SINE WAVES USED TO GENERATE GUST

FREQUENCY (rad/sec)	RELATIVE AMPLITUDE
0.3142	1.0
0.5027	1.0
0.7540	1.0
1.1938	1.0
1.9478	1.0
3.2044	1.0
5.1522	0.2
13.1319	0.1
19.9806	0.1

substantial increase in the controlled element lag. The end of retention is defined as the point in time where an abrupt change in manipulator motion begins, and it is most evident in transitions where the nonlinear control phase can be clearly seen.

The nonlinear control phase starts at the end of retention and continues until the system error has been reduced (approximately) to within the postfailure steady-state envelope. This envelope may exhibit a further, more gradual decay during the first seconds of this steady-state period, corresponding to an additional adjustment phase (ref. 1) wherein the pilot's final adaptation is being achieved by an optimizing process. Identifying the nonlinear control phase duration in the data is a subjective process and it is sometimes difficult. These phases are evident in the yaw rate errors of the present data.

TYPICAL RESULTS.—Typical examples of failures from configuration A are shown in figures 3 and 4. This prefailure configuration was designed to give enough damping that the effective airframe would be subjectively good and no pilot rudder control would be needed. Despite this, the pilot used the rudder intermittently before the failure, as shown in figures 3 and 4, presumably because of the high level of gust disturbance (which excited the aircraft) and the fact that the pilot knew a failure was coming. Interestingly enough, the approximate level of rudder gain that this represents is enough to make the postfailure configuration stable. Hence, the pilot appears to close a low gain rudder loop with configuration A in anticipation of the failure. With the soft-failures involving configuration A, a retention phase and a transition phase are both evident. The transition phase is defined as the interval following failure during which the pilot has not yet reduced the yaw rate amplitude envelope to the steady-state level (fig. 3). It shows up just after a failure as a period during which the yaw rate amplitude envelope builds up and then decreases. The retention phase is obvious in that during the first few seconds following failure the pilot continues to respond as he did with the prefailure configuration. It is interesting to note the relatively long duration of the retention phase (several seconds) for this situation, as shown in figure 3.

Typical failures from configuration B are shown in figures 5 and 6. This prefailure configuration was designed to give a lightly damped Dutch roll nuisance mode, but one that was not destabilized by aileron control of bank angle. The figures show that the pilot used larger rudder amplitudes with B than with A. A rapid recovery following failure can be seen in the yaw rate of figure 6.

Configuration C failures are typified by figures 7 and 8. This configuration was similar to B except that bank angle control with aileron tended to destabilize the Dutch roll mode because of the adverse $\omega_{\varphi}/\omega_d$ ratio. These figures show that the pilot's rudder activity before and after the failure is about the same. Rapid recovery following a hardover failure is shown in figure 8. On some of the runs involving soft-failures with configuration C it was not possible to detect the failure from the time traces of yaw rate, bank angle, aileron, and rudder because they all look essentially the same (in magnitude and frequency content) before and after failure. This suggests a rapid adaptation for such cases. An example is given in figure 7. The hardover failures for configuration C show a retention time of about 0.5 second. After retention a large rudder pulse (also lasting about 0.5 second) was applied in the direction to oppose the input, followed by a crude rudder step of opposite polarity lasting for several seconds. This opposite step was needed to cancel the yawing moment due to the aileron that the pilot was using to remove the bank angle induced by the failure transient, as shown in figure 8.

The data for configurations B, C, and E typically show a crossfeed by the pilot of aileron to rudder control. The effect of the crossfeed is to augment N'_{δ_a} , thereby reducing or cancelling the adverse effect of $\omega_{\varphi}/\omega_d$. Although configurations C and E are the only ones

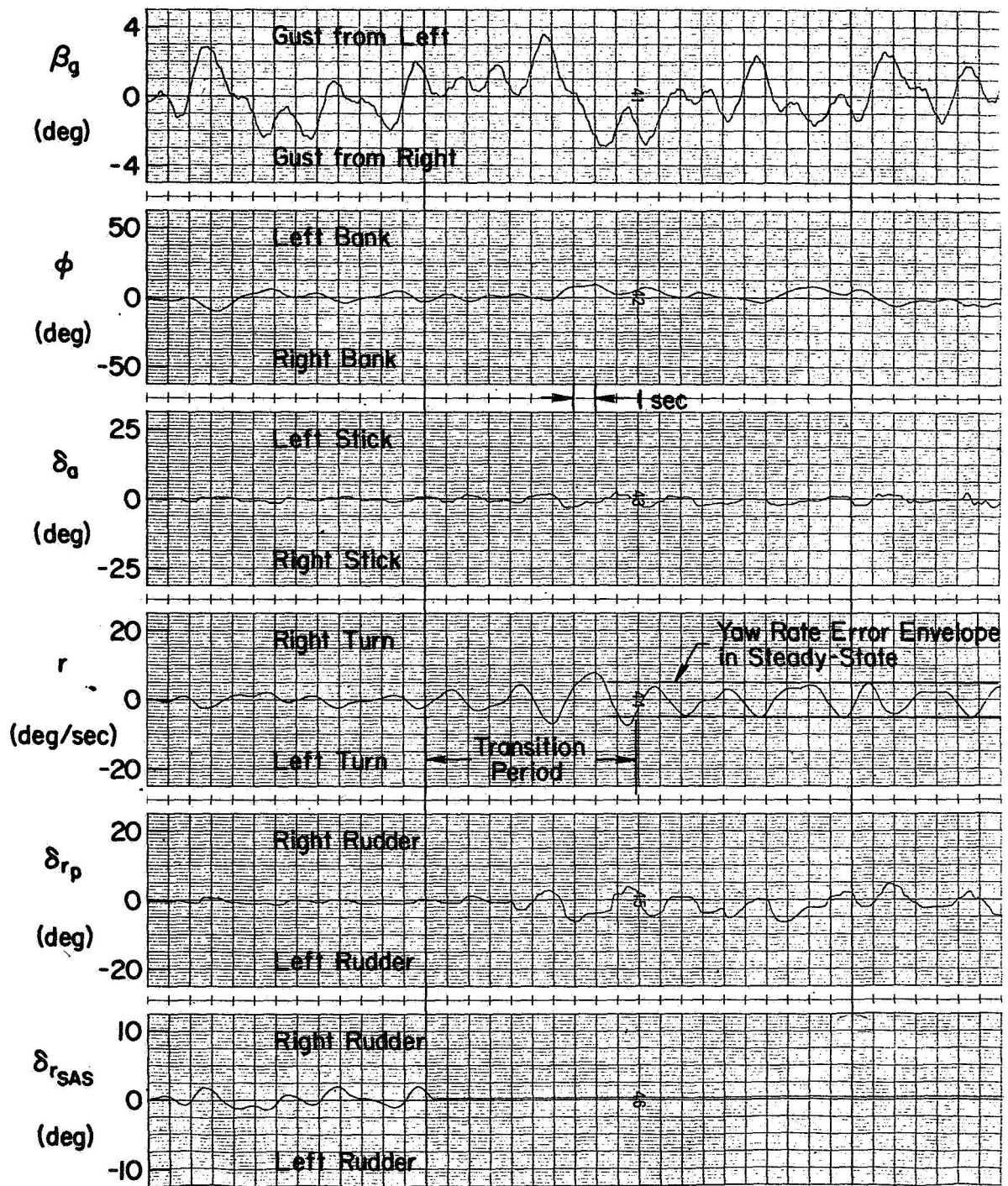


Figure 3. — Example of a soft-failure with configuration A.

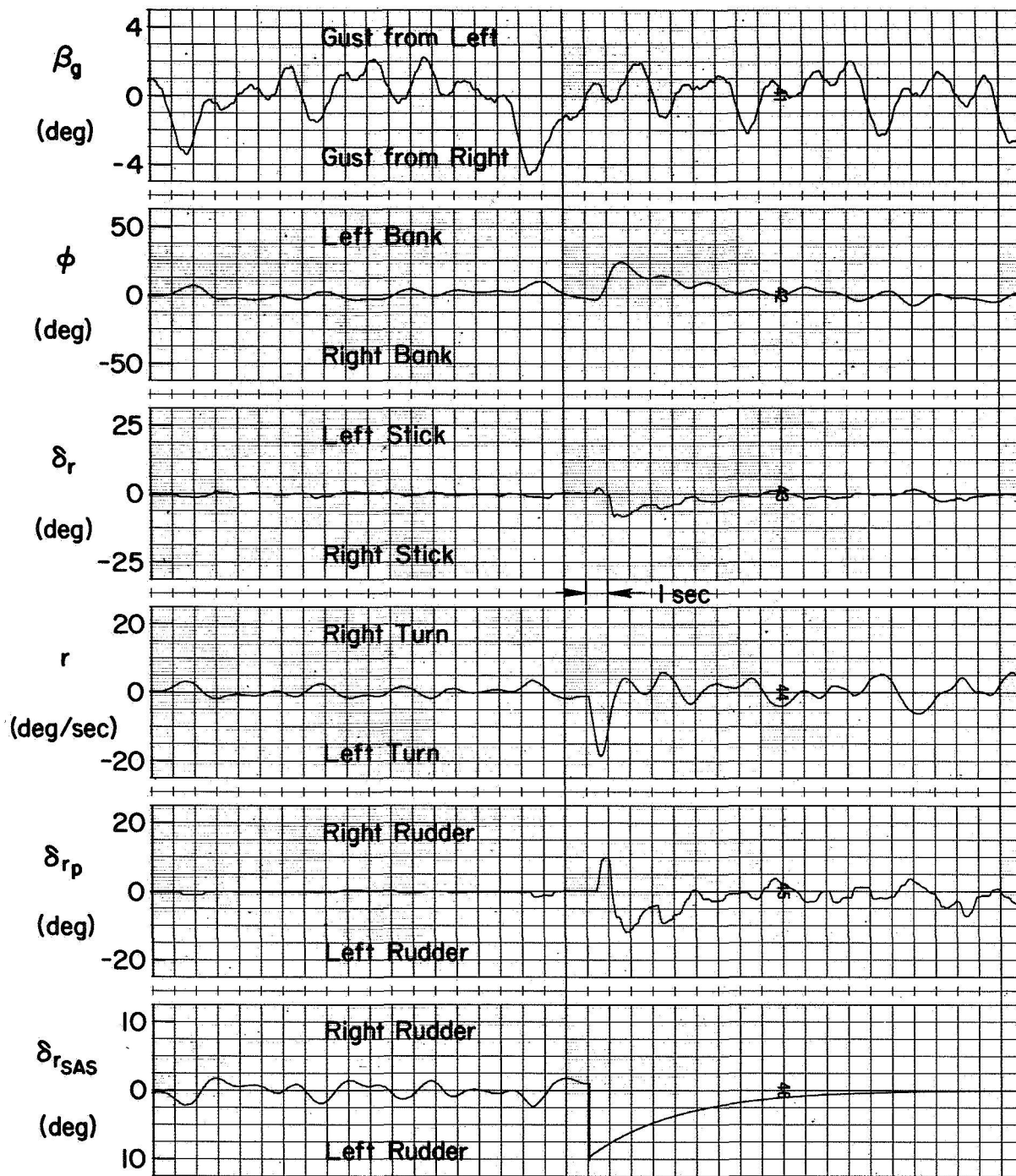


Figure 4. — Example of a hard-failure with configuration A.

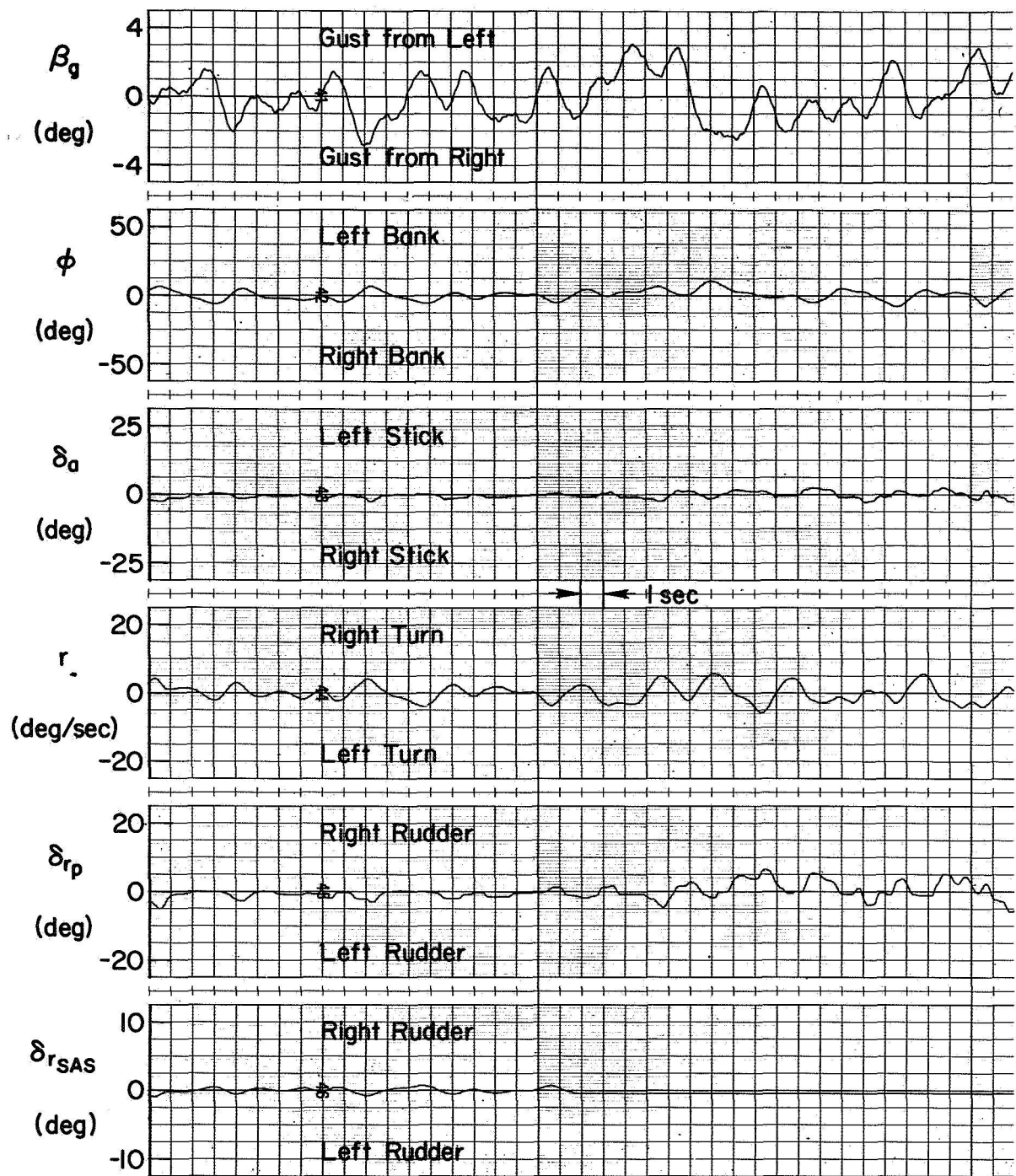


Figure 5. — Example of a soft-failure with configuration B.

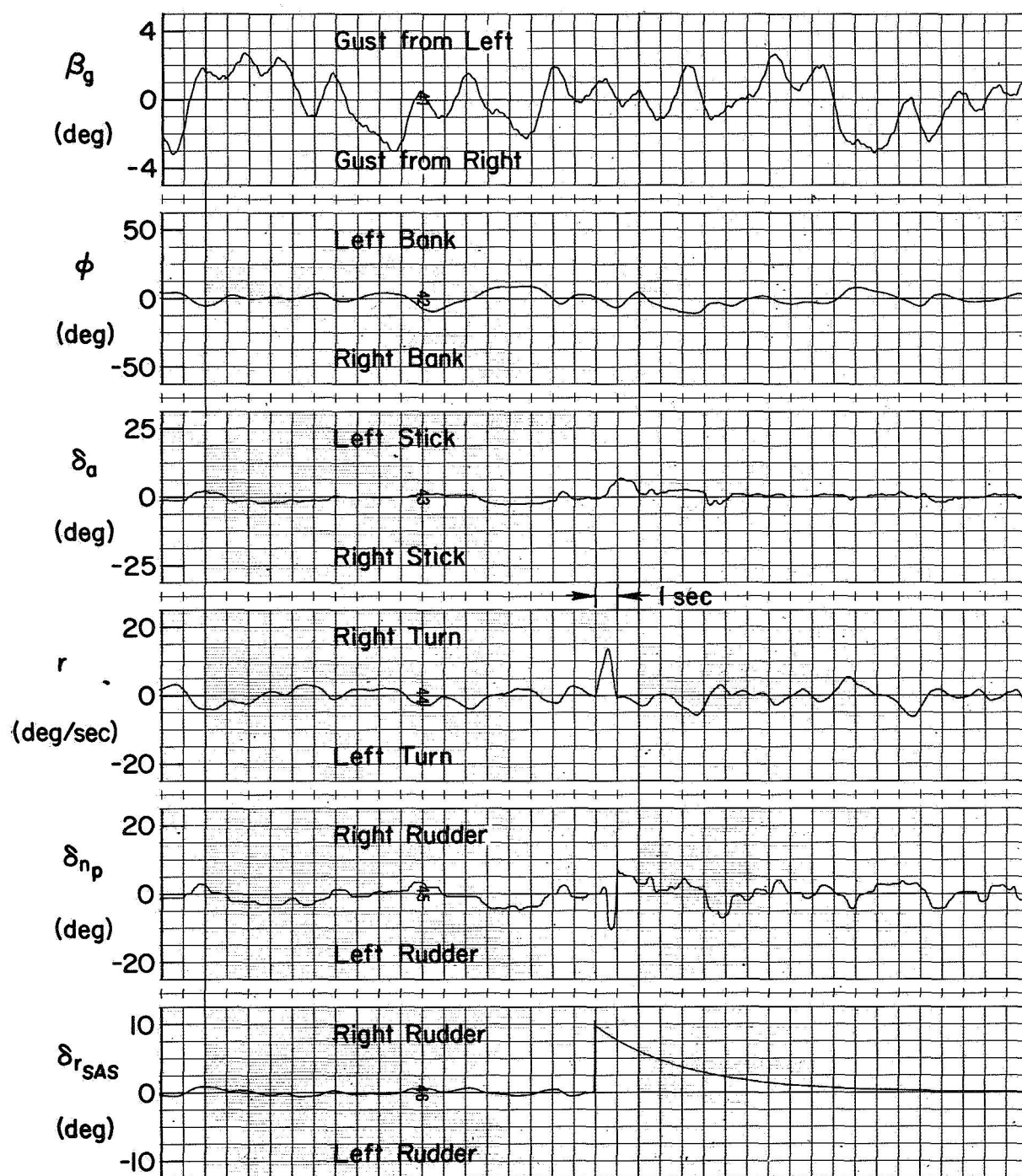


Figure 6. — Example of a hard-failure with configuration B.

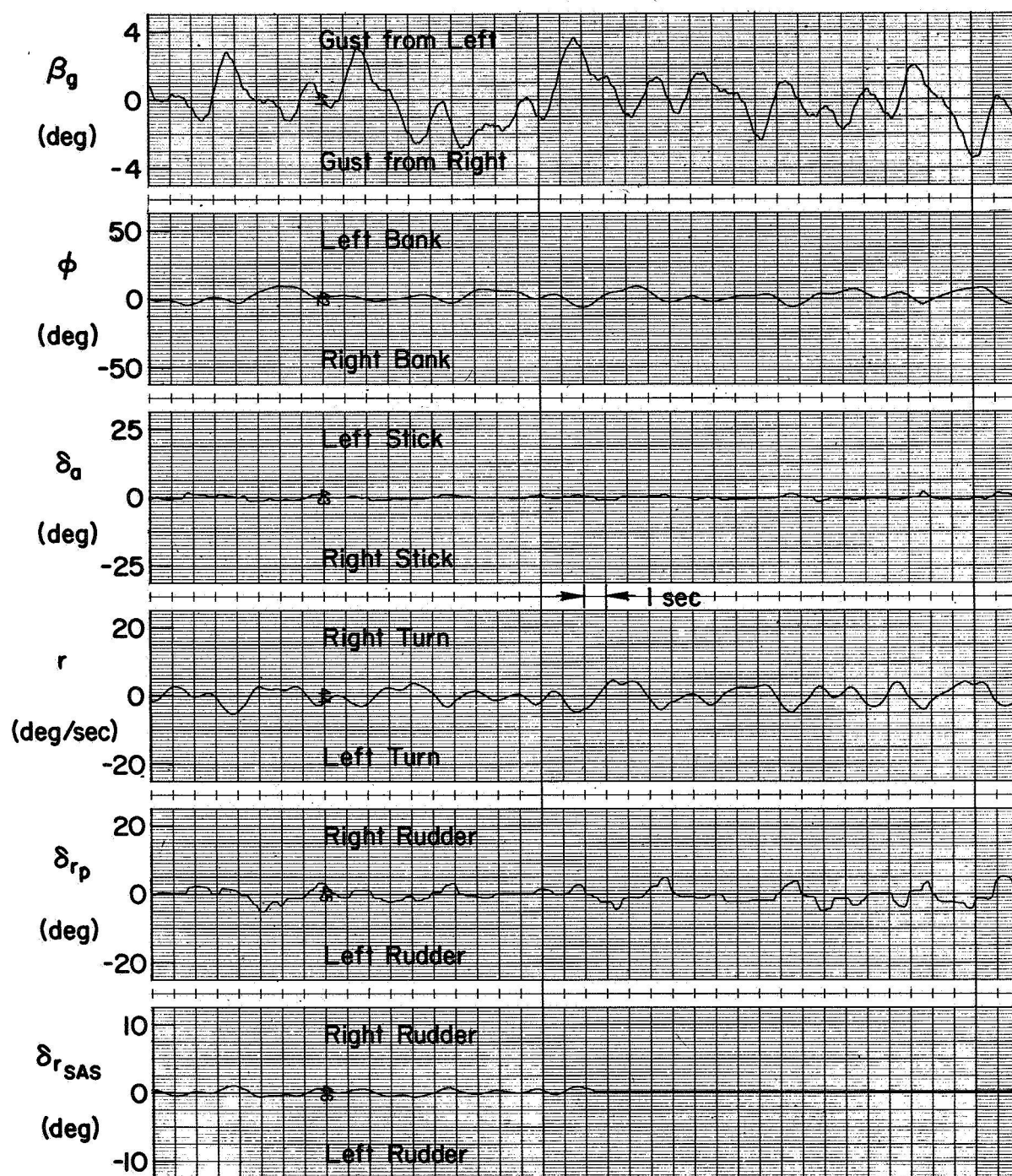


Figure 7. — Example of a soft-failure with configuration C.

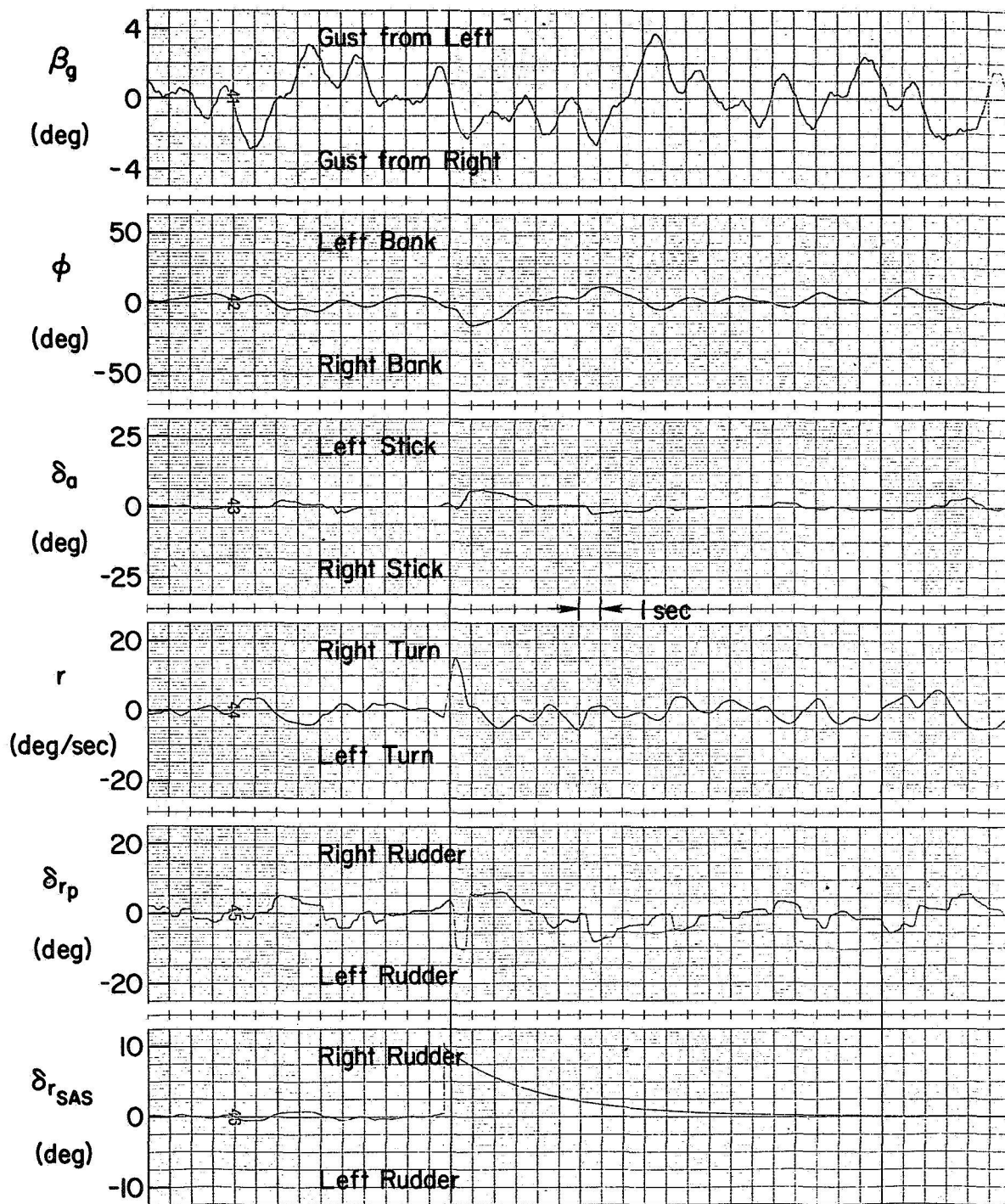


Figure 8.—Example of a hard-failure with configuration C.

requiring such a crossfeed, it was found that the pilot also used crossfeeds with A and B (which already exhibited pole/zero cancellation for the Dutch roll mode). This may be explained by the pilot learning that a crossfeed was desirable after the failure for better performance, and that a prefailure crossfeed did not reduce performance. These results support the graceful degradation hypothesis.

The data for all 90 runs were analyzed by measuring typical transition parameters and tabulating these for comparison of the configurations. The parameters included:

- Maximum yaw rate error peak during transition for soft-failures
- Magnitude of first yaw rate error peak for hard-failures
- Transition duration (or total settling time); the transition is considered to last as long as the yaw rate error peaks remain outside the envelope of the postfailure steady-state error peaks
- Number of significant yaw rate error overshoots during transition
- Magnitude of first rudder pulse for hard-failures
- Duration of first rudder pulse for hard-failures
- Pilot's retention duration for soft-failures, or his time delay prior to applying the initial rudder pulse for hard-failures

The tabulated results are summarized below.

The distribution of maximum yaw rate peaks for the soft-failures shows that the relative number of runs with yaw rate errors large enough to give an obvious transition region increases as the prefailure dynamics improve. The hard-failure data show that configurations B and C give smaller maximum yaw rates than A. Both of these results support the graceful degradation hypothesis.

The distribution of transition durations for soft-failures (which required that the error increase above the postfailure asymptotic level in order to be measured) showed that this only occurred with the good prefailure dynamics. Since the poor dynamics (configuration C) did not peak above that level, it strongly supports the graceful degradation hypothesis. The results are inconclusive for the hard-failures. Perhaps the duration is affected more by the hardover transient acting as a disturbance (the same in all cases) than it is by the change in dynamics.

The distribution of yaw rate error overshoots for the soft-failures showed that no significant error overshoots occurred for the poor prefailure dynamics while they occurred frequently with the good prefailure dynamics, again supporting the hypothesis. For the hard-failure data, the distribution of the number of yaw rate error overshoots does not lead to a conclusion regarding graceful degradation.

In summary, most of the soft-failure data supports the graceful degradation hypothesis of better performance being associated with smaller changes in the controlled element dynamics. This was also true for the maximum error peaks of the hard-failures. Some of the hard-failure data are inconclusive. There was no evidence to support a hypothesis contrary to that of graceful degradation being associated with smaller changes in the controlled element characteristics.

These results emphasize system performance and pilot adaptation following failure. A more subtle aspect of graceful degradation involves modification in prefailure pilot adaptation to obtain improved performance immediately following augments failure. Two types of prefailure pilot control action were observed which were not only not required by the configuration being flown at the time, but were slightly detrimental to the control of the prefailure configuration. They did result in improved control immediately after the augments failed. The first involved the pilot's using a small amount of rudder control with configuration A, and the second involved using an aileron-to-rudder crossfeed with configuration B. This pilot response phenomenon has been dubbed "conditional adaptation,"

because it is conditioned or influenced by a future event that is defined probabilistically. There is considerable evidence of conditional adaptation in the data, and further research is needed to relate the adaptation to the probability and severity of failure. A similar control strategy is noted in reference 7.

AVERAGE RESULTS.—Digital analyses were made for each type of failure to obtain average results as a function of time. The tape recorded data were sampled at 0.1-second intervals and computations were made from 10 seconds before the failure to 20 seconds after the failure. Five runs were selected for analysis for hard- and soft-failures with each of the three configurations (30 runs in all). Each set of five runs formed an ensemble, and averages were made within each ensemble to obtain

- (1) time-varying mean (at 0.1 second intervals)
- (2) time-varying standard deviation (at 0.1 second intervals)
- (3) time-varying autocorrelation function (at 1 second intervals)

These calculations produce average results that show gross trends, but they obscure fine detail in the data. Hence they complement (rather than replace) the preceding discussion, which in effect considers typical properties of the fine detail.

The yaw rate standard deviation (σ_r) is most pertinent to the question of graceful degradation. For soft-failures the mean value of yaw rate will approach zero as the number of runs in the ensemble increases. However, the standard deviation from the mean is an indication of the yaw rate within a given run. In particular, differences in the magnitude and duration of variability from the mean just after failure for different configurations would be an indication of relative degradation of performance (increased error).

Figure 9 shows the time variation through failure of the yaw rate standard deviation for soft- and hard-failures with configurations A and C.

Figures 9(a) and (c) are plots of 3-second averages of σ_r for the period from 10 seconds before a soft-failure to 20 seconds after the failure. Figure 9(a) shows that with configuration A there is a region of yaw rate just after a soft-failure which is higher than that in either the prefailure or postfailure steady state. This indicates a period of degraded performance. Figure 9(c) shows that such a situation does not occur with configuration C (for soft-failures) where the yaw rate appears to increase slowly and monotonically to the postfailure steady-state level. This result is consistent with the typical results cited above where no transition region could be identified for soft-failures from configuration C.

For hard-failures, both configurations A and C show a transient increase in σ_r following failure. This is a result of the hardover transient causing a large error to occur regardless of the change in the dynamics. It suggests that (for hardovers) graceful degradation relates more to the postfailure dynamics being bad than to any relative differences in prefailure and postfailure dynamics.

PILOT RESPONSE MODEL

A composite pilot model for step changes in controlled element dynamics results from the available single- and multiple-loop data, which applies to skilled subjects in practiced situations. The salient features of the model can be expressed in terms of a single-loop tracking task. It is applicable to the failed loop in a multiple-loop situation if only one loop involves a significant change in the controlled element dynamics.

The composite model is quite like the earlier single-loop model of reference 1 which divides the transition into phases with respect to time, each having certain distinguishing characteristics:

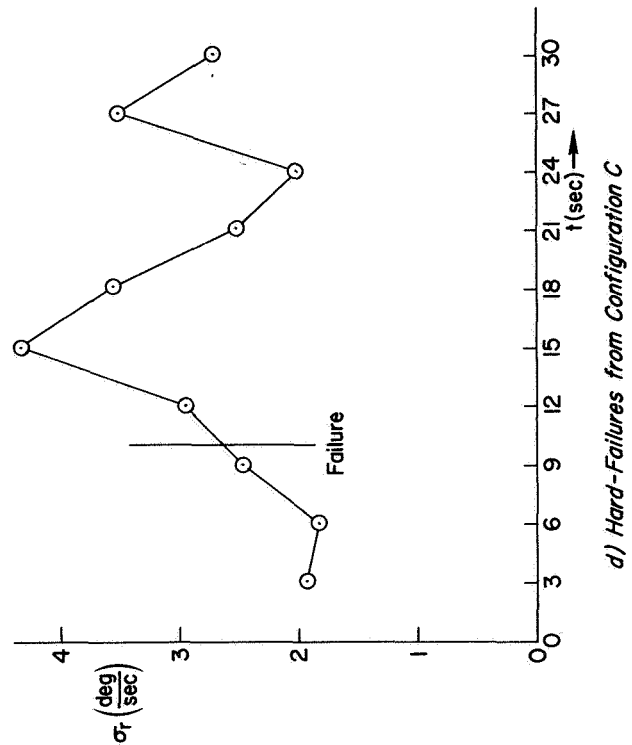
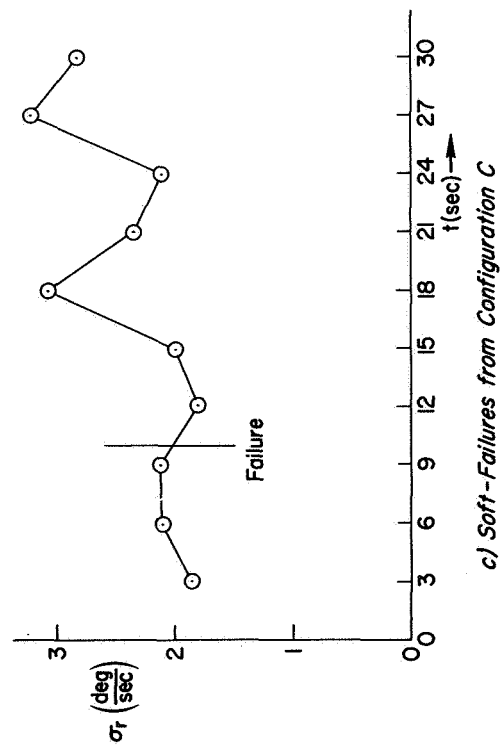
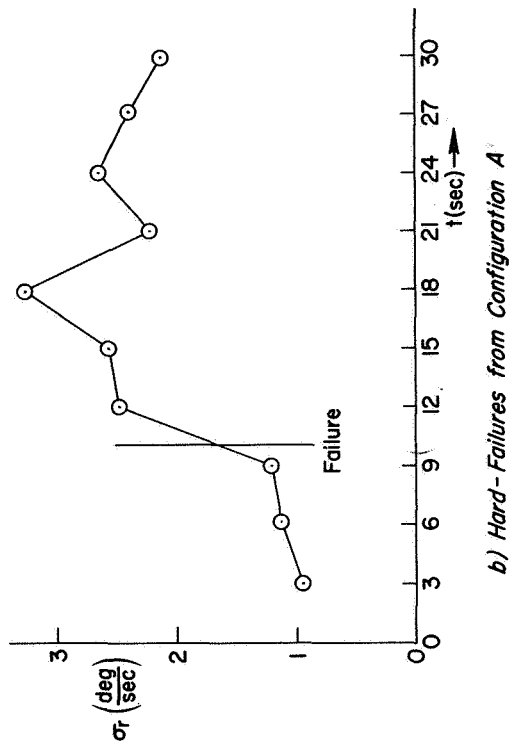
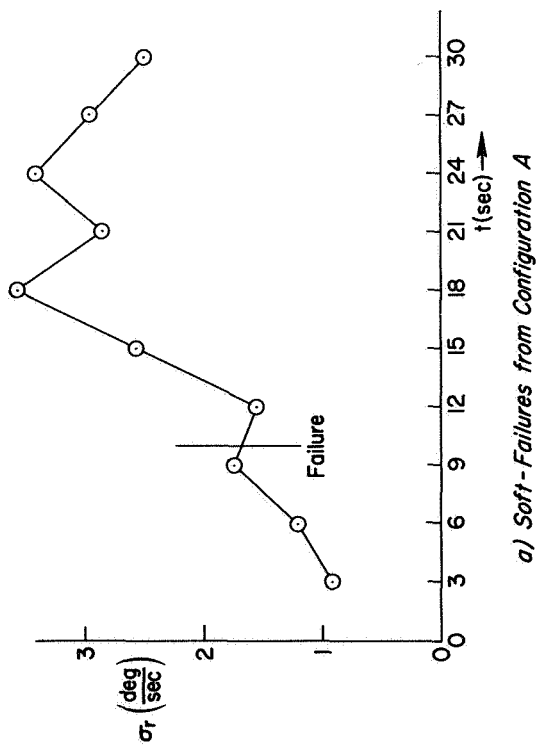


Figure 9. --Yaw rate standard deviation.

(1) Prefailure steady-state; stationary compensatory tracking, with the possible use of conditional adaptation.

(2) Retention; prefailure pilot describing function plus the postfailure controlled element dynamics operating closed-loop.

(3) Nonlinear control; large control actions (sometimes time-optimal) which stabilize the system and reduce the error to some acceptable level.

(4) Adjustment phase; pilot describing function has same form as that for postfailure steady-state phase, but parameter values are being adjusted to minimize system error.

(5) Postfailure steady-state; stationary compensatory tracking.

The retention phase starts at the failure. Its end is defined as the point in time when an abrupt change in pilot control output begins. It is most evident in transitions where the nonlinear control phase can be clearly seen.

The nonlinear control phase (if present) starts at the end of retention and continues until the pilot control output has been reduced (approximately) to within the postfailure steady-state envelope. Time-optimal control with fixed-bang amplitudes is an appropriate limiting case; as shown in reference 1. The nonlinear control phase may include a learned response that is merely triggered by the failure (such as a feedforward step). For soft-failures with some systems the nonlinear phase is not present. It is present only when required, i. e., when linear control is inadequate to maintain reasonable system errors.

The postfailure steady-state phase starts at the end of the nonlinear control phase, to a first approximation. In some cases, the envelope of system error decays gradually following the nonlinear control phase. This corresponds to an adjustment phase wherein the pilot's steady-state adaptation is being achieved by a relatively long term (e. g., 5 to 10 seconds) optimization process.

Block diagrams of the various phases of the model are given in figure 10. The nonlinear phase can be described by a nonlinear controller alone in the case where the retention phase is highly unstable. A linear controller plus feedforward is appropriate for this phase in the hardover cases with a stable retention phase (e. g., configuration C), in which instance the signal fed forward should resemble the hardover transient. A nonlinear controller plus feedforward may appear appropriate for hardover case with unstable retention, but the response is probably indistinguishable from the nonlinear-controller-alone version. Whichever model is easier to use is appropriate.

The results show that the general form of model given in reference 1 is still valid with a few extensions and modifications:

(1) Numerical estimates are now available for the retention duration, when this phase exists.

(2) The pilot compensates for failure transients (hardovers, etc.) during the nonlinear control and postfailure steady-state phases. In some cases this appears to be a feedforward signal which cancels the failure transients, although this was not apparent for the most difficult postfailure controlled elements. An alternative is to treat the transient as an initial condition to nonlinear control and then produce a trim bias in postfailure steady state.

(3) Time-optimal control alone during the nonlinear phase is still a valid idealization for the soft-failures, as an attainable limiting case. However, some suboptimal control mode is more typical.

(4) Postfailure steady-state error characteristics were generally larger than that predicted by an optimized quasi-linear model. This suggests that any adjustment (optimization) phase is relatively long term and does not end immediately following the nonlinear phase.

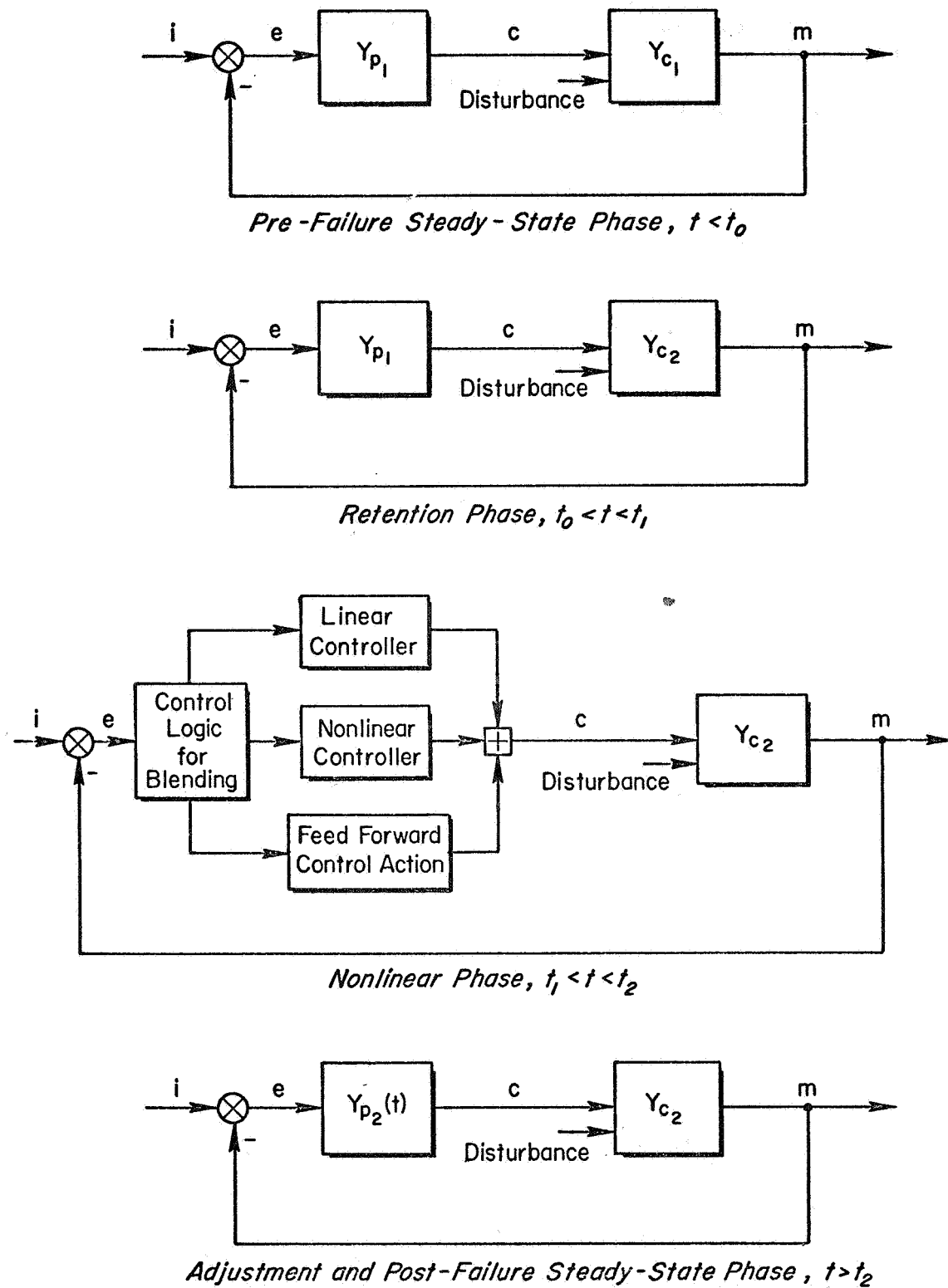


Figure 10. —General pilot model for step change in controlled element dynamics.

(5) Prefailure conditional adaptation may be present. This means that the pilot may use an adaptation prior to failure which is not a steady-state optimum but which will improve the system performance immediately after failure (such as a modified gain or a crossfeed).

IMPLICATIONS FOR DESIGN

The experimental results lead to several considerations pertinent to flight control system design. These are summarized in the following paragraphs.

The graceful degradation hypothesis is supported by the experimental results. Therefore, a flight control system design that gives less change in effective controlled element following a failure will give better transition performance.

The transition performance with hardover failure was relatively invariant with the amount of change in the dynamics at failure. The performance with soft-failures tended to vary as a function of the change in dynamics, but it was generally better than that for hardovers. The relative invariance of the hardover failures is a consequence of the retention times being about the same due to alerting. It suggests that if alerting can be designed into (or accompany) soft-failures, performance will be improved.

For soft-failures the crux of the entire graceful degradation situation lies in the closed-loop system stability during the retention phase; that is, the stability of the prefailure pilot model Y_{p_1} and controlling the postfailure vehicle Y_{c_2} . If $Y_{p_1}Y_{c_2}$ is stable, closed-loop, then a critical situation generally does not develop after a failure.

A flight control (e.g., stability augmentation) system should be designed to keep the pilot in the loop prior to a failure rather than to allow (or require) him to merely monitor the controller/vehicle system activity. This might be accomplished by giving the pilot the task of controlling a simultaneous model of the actual system. One payoff for this is obvious—the pilot will have a shorter retention period. For example, a pilot who has his feet on the pedals at the time of failure because he is "in the loop" will have a shorter time delay prior to initial rudder input than a pilot who is merely an observer and has his feet on the floor. For some kinds of failure a long retention period or delay would be catastrophic.

A pilot tends to use conditional adaptation, which means that his prefailure control activity is a function of the probability of failure and the severity of the transition. If $Y_{p_1}Y_{c_2}$ closed-loop (or Y_{c_2} alone) presents difficult control characteristics, then the pilot will anticipate a failure by modifying his prefailure adaptation Y_{p_1} to improve $Y_{p_1}Y_{c_2}$. This will be something other than the stationary values appropriate for Y_{c_1} . Consequently, a control system should not only make prefailure flight control available, but it should be designed so that considerably improved $Y_{p_1}Y_{c_2}$ stability is possible with simple modifications in prefailure control activity (as was the case in the multiloop experiments).

The data have implications for pilot training. During training of the pilot subjects in the experimental tasks, performance following a failure improved rapidly as the number of practice runs increased. Even hardover failures gave little performance degradation when encountered regularly. Any failure can be difficult if it represents a significant change from recent experience (as was found during early failure practice runs following only steady-state control of each configuration), which implies that pilots should have frequent refresher training to keep them current with regard to flight control system failure situations.

CONCLUSIONS

The experimental results indicate that the graceful degradation hypothesis is valid. Thus an improvement in tracking performance can be expected during a transition in controlled element dynamics if the change in dynamics is diminished.

A previous model for human pilot dynamic response during controlled element transitions (ref. 1) was verified and extended. The verification shows that the following transition response phases do exist in general:

- (1) Prefailure steady-state
- (2) Retention
- (3) Nonlinear control
- (4) Postfailure steady-state

The extensions relate to the existence of hardover (deterministic) signals in the system and occur in the nonlinear control and postfailure steady-state phases.

REFERENCES

1. Weir, David H.; and Phatak, Anil V.: Model of Human Operator Response to Step Transitions in Controlled Element Dynamics. NASA CR-671, Jan. 1967.
2. Sheridan, Thomas B.: Time-Variable Dynamics of Human Operator Systems. AFCRC-TN-60-169, Mar. 1960.
3. Sadoff, Melvin: A Study of a Pilot's Ability to Control During Simulated Stability Augmentation System Failures. NASA TN D-1552, Nov. 1962.
4. Young, Laurence R.; Green, David M.; Elkind, Jerome I.; and Kelly, Jennifer A.: The Adaptive Dynamic Response Characteristics of the Human Operator in Simple Manual Control. NASA TN D-2255, Apr. 1964.
5. Elkind, Jerome I.; and Miller, Duncan C.: Adaptive Characteristics of the Human Controller of Dynamic Systems. AFFDL-TR-66-60, July 1966.
6. Elkind, Jerome I.; and Miller, Duncan C.: On the Process of Adaptation by the Human Controller. Paper presented at Third Congress of the International Federation of Automatic Control (London), 20-25 June 1966.
7. Swarder, D. D.: On the Control of Stochastic Systems. Univ. of Southern California USCEE Rept. 217, July 1967.

20. Model of the Adaptive Behavior of the Human Operator in Response to a Sudden Change in the Control Situation*

*A. V. Phatak and G. A. Bekey
University of Southern California*

An adaptive model is presented to describe the behavior of the human operator in response to sudden changes in plant dynamics and transient disturbances. The plant simulated for tracking experiments is approximately second order and has rate and attitude feedback augmentation for increased stability. The failure of the rate sensor and/or the attitude sensor results in a sudden transition in the order and gain of the effective plant dynamics. These failures may be accompanied by hard-over transient conditions in either the rate or attitude sensors. The adaptive model suggested has a variable structure, contains mode-switching based on pattern recognition as evidence, and incorporates the decision-control logic required for successful adaptation to failures. The model in effect attempts to mimic the control strategy or algorithm used by a trained operator.

INTRODUCTION AND BACKGROUND

The ability of the human operator to adapt to a changing environment has received much attention of late. Specifically, these efforts have mainly been in modeling the adaptive response of the human operator in a compensatory or pursuit tracking task when faced with a time-varying control situation. (refs. 1 to 9) The different types of time-varying control situation possible are

- (1) A change in the input characteristics, e.g., statistical or transient variations.
- (2) A change in the display, e.g., compensatory to pursuit, or vice-versa, and preview variations.
- (3) A change in the effective dynamics of the plant being controlled.

The above three categories are not necessarily mutually exclusive, and it is very likely that in some cases the type of change may be indistinguishable. For example, a time-varying plant gain may be interpreted as a modulated input (refs. 10 and 11), the modulation being a function of the gain variation; also, deterministic transients accompanying plant augmentation

*This research was supported in part by the National Aeronautics and Space Administration under NASA Grant No. NGR 05-018-022.

failure (a decayed exponential or a step) can be considered equivalently as transient inputs to the system. Thus, it is not so much the type of change that is important, but, rather, the effect on the control situation caused by it.

PROBLEM STATEMENT.—The purpose of this paper is to study and model the adaptive behavior of the human operator in response to a type C time-varying control situation. Specifically, the change in the effective dynamics of the plant is assumed to be sudden with a finite number of posttransition dynamics allowed. The change in the plant dynamics is so chosen as to result in overall system instability unless the human operator modifies his structure to that required for the new dynamics. There have been a number of studies in the past (refs. 8 and 9) which have attempted to model human-operator adaptive behavior in response to a sudden change between stationary control situations. The principal idea in these models has been the partitioning of the adaptive response into a finite number of distinct phases and presentation of the corresponding human-operator models. The need for decision elements in adaptive models of human operators has been recognized (refs. 5, 6, and 8), but no attempts have been made to identify the complete decision-control logic. Most of the studies assume an explicit identification phase where the operator recognizes the posttransition plant dynamics using a model-matching scheme or a decision-theory model. It is obvious that mathematical models based on such methods become quite complicated as the number of posttransition plant dynamics increases. The operator is assumed to be well-trained in adapting to the various alternative, but finite, types of failure situations. Consequently, the operator must have evolved or learned a quasi-deterministic algorithm for successful adaptation to the different types of failures. This argument is similar to the one given for using quasi-linear describing functions as human-operator models for stationary steady-state tracking. Any slight variation from this algorithm may be considered as a second-order effect and can heuristically be explained as noise in the decision logic used by the algorithm or as measurement noise in the observed state variables.

A first step would be to identify the complete on-line algorithm which models the decision-control logic in human-operator adaptation to failures. This paper attempts to fill this need. The proposed model incorporates sequential mode-switching between a finite number of adaptation structures based on sequential identification of control situations using pattern recognition in phase space.

THE EXPERIMENTAL SITUATION

A practical example of a time-varying control situation is an aircraft whose stability augmentation may fail suddenly. Data for such a situation was made available by the courtesy of Systems Technology Inc., Hawthorne, California. The data were taken in a series of experiments where the human operator was instructed to track (keep wings level) a single-axis roll-control task of a VTOL aircraft in hover, as shown in figure 1 (ref. 12). The airframe is assumed to be pure inertia and has rate and attitude feedback augmentation for increased stability. The input is low frequency and random, appearing with a cutoff frequency of 1.5 rad/sec (sum of six sine waves with a root mean square value on the operator's display of 5°). For complete details on the component dynamics and the mechanization of the single-axis experiments, see reference 12.

A change in the effective plant dynamics is induced by failing either or both of the feedback loops. At failure, the outputs of the rate and attitude sensors either go to zero or have step or ramp transients to a nonzero constant magnitude bias. The above three kinds of sensor failure are referred to as soft, hard, and ramp failures, respectively. The sequence

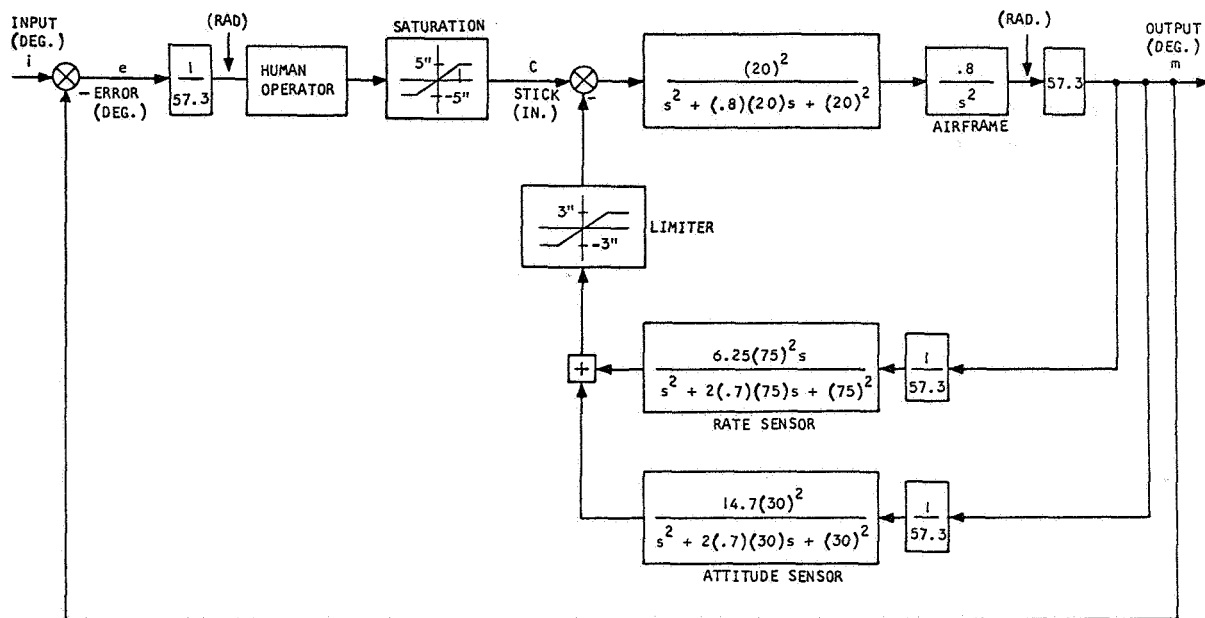


Figure 1.—Human operator in roll-tracking task.

TABLE 1.—TIME-VARYING CONTROL SITUATIONS STUDIED

SITUATION	PROBABILITY OF OCCURENCE
Both loops fail to zero FB	24%
Rate loop fails hardover	14%
Attitude loop fails hardover	32%
Attitude loop fails ramp	22%
Rate loop fails to zero FB	4%
Attitude loop fails to zero FB	4%

of failures presented to the human operator was randomized in accordance with reliability and failure mode data on system components. Table 1 presents the types of changes in effective dynamics (with or without failure transients) encountered by the human operator along with their likelihood or occurrence.

In short, transitions proposed in table 1 occur among the corresponding prefailure and postfailure effective plant dynamics shown in table 2. Hereafter, the four augmentation levels are referred to by letters A, B, C, and D, respectively.

Two separate sets of transitions were explored experimentally as shown in table 3. Set 1 experiments had three distinct postfailure plant dynamics and six different failure situations. Set 2, however, had one fixed postfailure plant and only two failure situations. The two sets of experiments were done in separate experimental sessions with no mixing in any one session.

TABLE 2. — EFFECTIVE PLANT DYNAMICS

Augmentation Level	Effective Plant Dynamics
Rate and Attitude A	$\frac{1.72}{[s^2 + 2(.6)(5.1)s + (5.1)^2]}$
Rate Only B	$\frac{320}{s(s+18.6) [s^2 + 2(.66)(10.4)s + (10.4)^2]}$
None C	$\frac{.8}{s^2}$
Attitude Only D	$\frac{.7154}{[s^2 - 2(.2)(3.3)s + (3.3)^2]}$

TABLE 3. — SETS OF TRANSITIONS SIMULATED

Pre-Failure Augmentation Level	Types of Failure	Number of Experimental Runs	Post-Failure Augmentation Level
Set 1 Rate Plus Attitude A	soft	4	Rate Only B
	ramp	22	
	hard	32	
	soft	4	Attitude Only D
	hard	14	
	soft	24	None C
Set 2 Rate Only B	soft	43	None C
	hard	17	

A typical experimental run lasted 3 to 4 minutes with the failure occurring at a random time, 1 to 3 minutes from the start. The subject was a well-trained pilot and was given 10 hours of training (180 transitions) in controlling various system failures prior to actual experimental runs. The pilot was not given any details about the system dynamics, failure rates, etc. which would tend to bias his control strategy. Training was the only source of learning available to the operator.

BASIC FEATURES AND CONSTRAINTS FOR HUMAN OPERATOR ADAPTATION

Before proceeding with the development of an adaptive model, it is imperative to know the steady-state models (describing functions) of the human operator for prefailure and post-failure plant dynamics. For each of the four levels of augmentation, it is possible to predict from existing data (ref. 13), the steady-state quasi-linear describing functions. They fit a four-parameter model of the form

$$Y_{H \cdot O}(j\omega) = KP \frac{(j\omega + ZP)}{(j\omega + PP)} e^{-(TAU)j\omega} \quad (1)$$

Estimated values for the parameters are as shown in table 4.

TABLE 4. — ESTIMATED HUMAN OPERATOR STEADY-STATE
MODEL PARAMETERS

Augmentation Level	KP	TAU	ZP	PP
A. Rate Plus Attitude	8.0	0.40	3.0	0.05
B. Rate Only	17.2	0.24	ZP = PP	
C. None	3.8 x PP	0.40	0.20	> 40
D. Attitude Only	6.0 x PP	0.10	0.20	> 40

On comparing tables 2 and 4, it is clear that a decrease in the level of augmentation results in an increase in the order of the effective plant dynamics and a corresponding reverse trend in the human-operator compensation from lag-lead to lead.

Any adaptive model of the human operator must satisfy the boundary constraints of prefailure and postfailure steady-state operator models. This implies that following failure the human-operator model must detect the occurrence of a failure and proceed with the modification required from its prefailure structure to postfailure steady-state requirements. Obviously the modification must be based on some kind of identification of postfailure plant dynamics, either explicit or implicit. Based on this inductive reasoning, the adaptive model must include the following phases and characteristics:

- (1) Prefailure steady-state operator model
- (2) Detection of failure or change in effective vehicle dynamics
- (3) Identification and modification
- (4) Postfailure steady-state operator model

Note that the phase following failure and prior to failure detection has been termed retention phase in the literature (ref. 9).

The model must incorporate the decision processes involved in going sequentially through the four phases. The overall adaptation must be rapid enough to avoid loss of control and ought to stabilize the system within a few seconds since the overall system is unstable in the retention phase.

Therefore, an adaptive model incorporating a higher level controller, called a supervisor (or monitor), capable of decision logic and able to reorganize the steady-state controller structure, seems appropriate. A block diagram of the same is presented in figure 2. It is the objective of this paper to give a mathematical description to the decision processes involving detection, identification, and modification used by the supervisor for successful adaptation, and this is given next.

PROPOSED ADAPTIVE ALGORITHM FOR SUPERVISORY CONTROL

The basic constraints and features of human-operator adaptation discussed in the previous section lead one to hypothesize that the supervisory-control algorithm is a sequential or multistage discrete decision process possessing the Markovian property (ref. 14). This is essential because decisions cannot wait until the complete time process is observed, but instead must be generated as the process unfolds in time. The assumption also permits one to replace the complexity of multihypothesis decisionmaking by sequential binary-hypothesis testing, which results in shorter decision times and rapid adaptation.

On the basis of the above arguments, a model for the supervisory control algorithm for the experimental situation in set 1 is proposed as in figure 3. The model for set 2 conditions

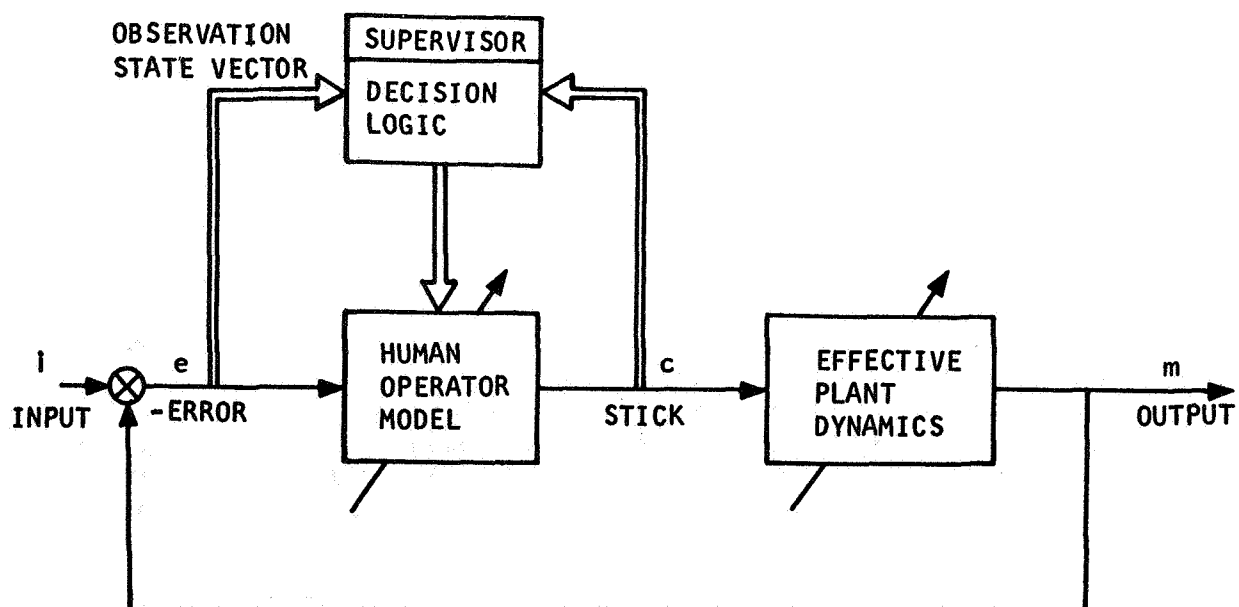


Figure 2. — Basic structure of human operator adaptive model.

follows by way of analogy and is presented in figure 4. Only the model for set 1 conditions is discussed in detail, leaving the analysis of the simpler situation in set 2 as a mere extension of arguments used in justifying the set 1 results.

The model of figure 3 includes decision and modification elements in a sequential order. Note that the modifications in the operator model lead in discrete steps to the new structure. The parameters on switching are allowed to vary from one trial to the next

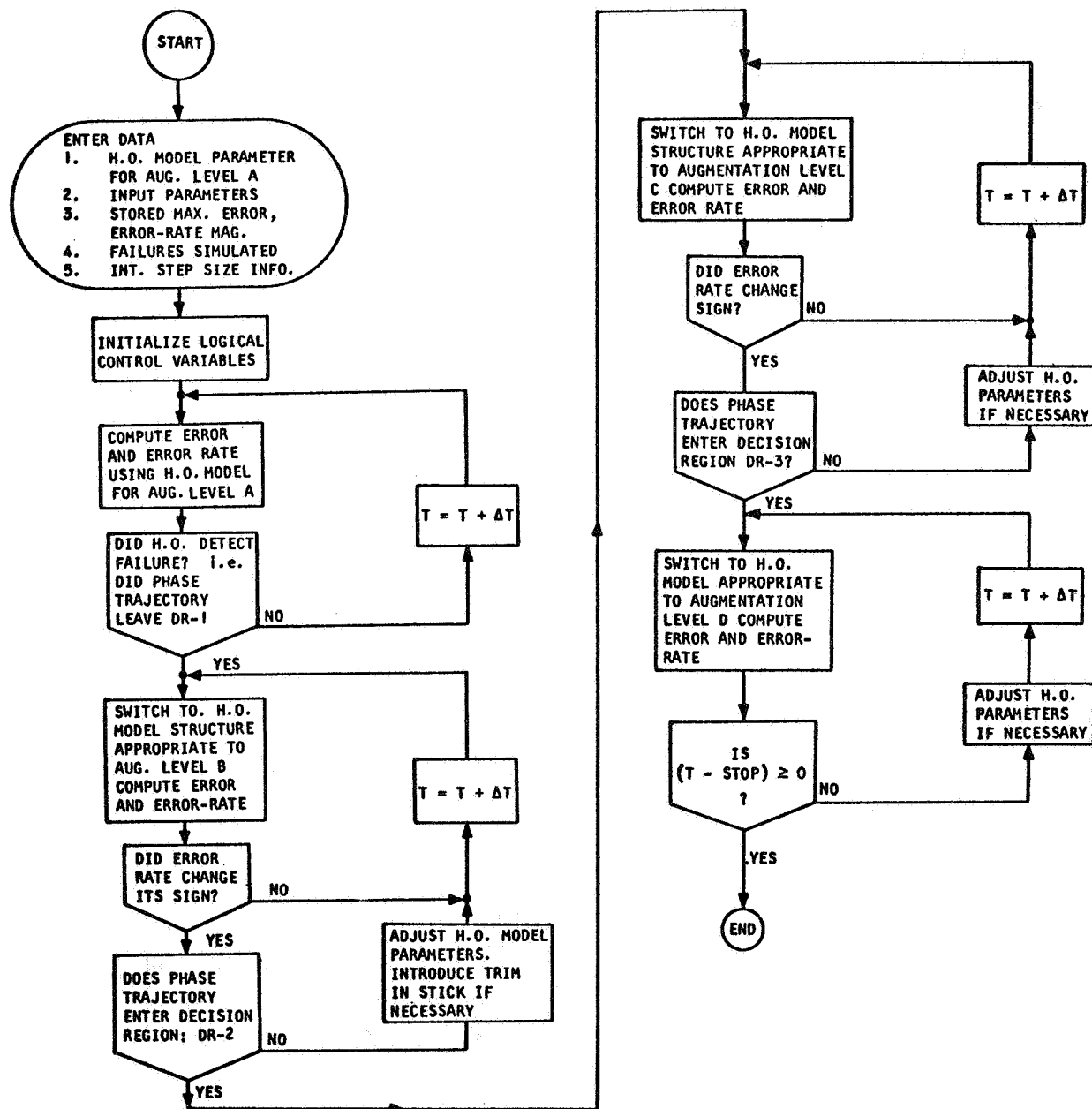


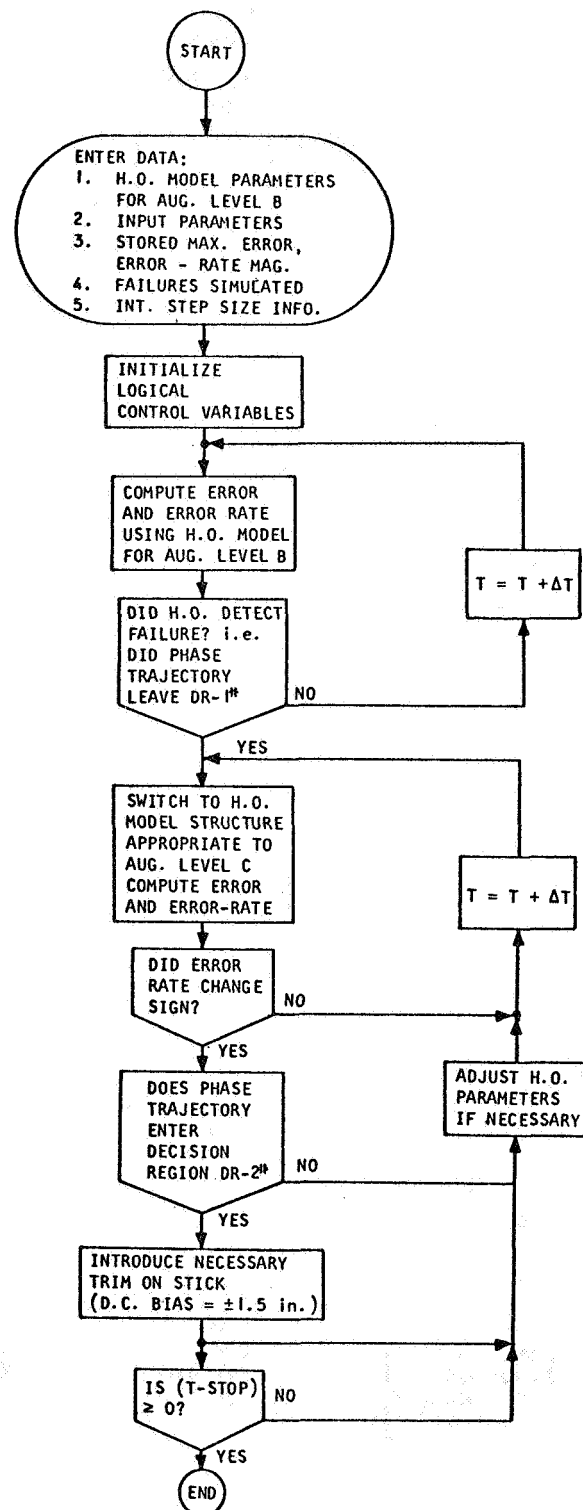
Figure 3. — Flow chart of supervisory control algorithm for set 1 experimental situation.

according to some suitable statistical distribution. This feature can provide the model with the ability to display run-to-run variability in response and hence make the model inherently stochastic. The range of parameter variations, however, must fall within the stability boundaries.

Analysis of experimental transition data provides the information necessary to identify the decision elements in figure 3, specifically, the significance of decision regions DR-1, 2, and 3, respectively, used for the purpose of partitioning the phase plane as shown in figure 5. Human-operator response data for two typical failures is shown in figures 6 and 7. Figure 6 shows the response to a soft failure in the augmentation from level A to level C, while figure 7 gives that for a hard failure in the augmentation from level A to level B. Corresponding phase-plane plots (error versus error rate) are presented in figures 8 and 9 respectively.¹ The decision regions DR-1, 2, and 3 which have yet to be defined are superimposed in figures 8 and 9 for later reference.

Note: TF=time of failure.

Based on available transition data (including that in figs. 6 through 9) the detection of failure seems to be dependent upon the range of values spanned by the observed-state variables, error and error rate, prior to failure. Here it is assumed that error rate is a directly perceived variable in addition to error. Since the overall system is unstable in the retention phase, either or both of the observed state variables must eventually exceed the prefailure bounds by some scale factor. This is indeed apparent in the phase-plane data of figures 8 and 9, where the error rate after failure suddenly increases and the phase trajectory leaves the decision region DR-1. It is postulated here that the human operator tracks and stores in memory the maximum absolute values of error and error rate for the prefailure steady-state conditions. If the observed variables exceed a scale factor of twice these stored upper



¹In phase-plane plots data is plotted every 0.2 seconds, and the points joined by a line. Figure 4. —Flow chart of supervisory control algorithm for set 2.

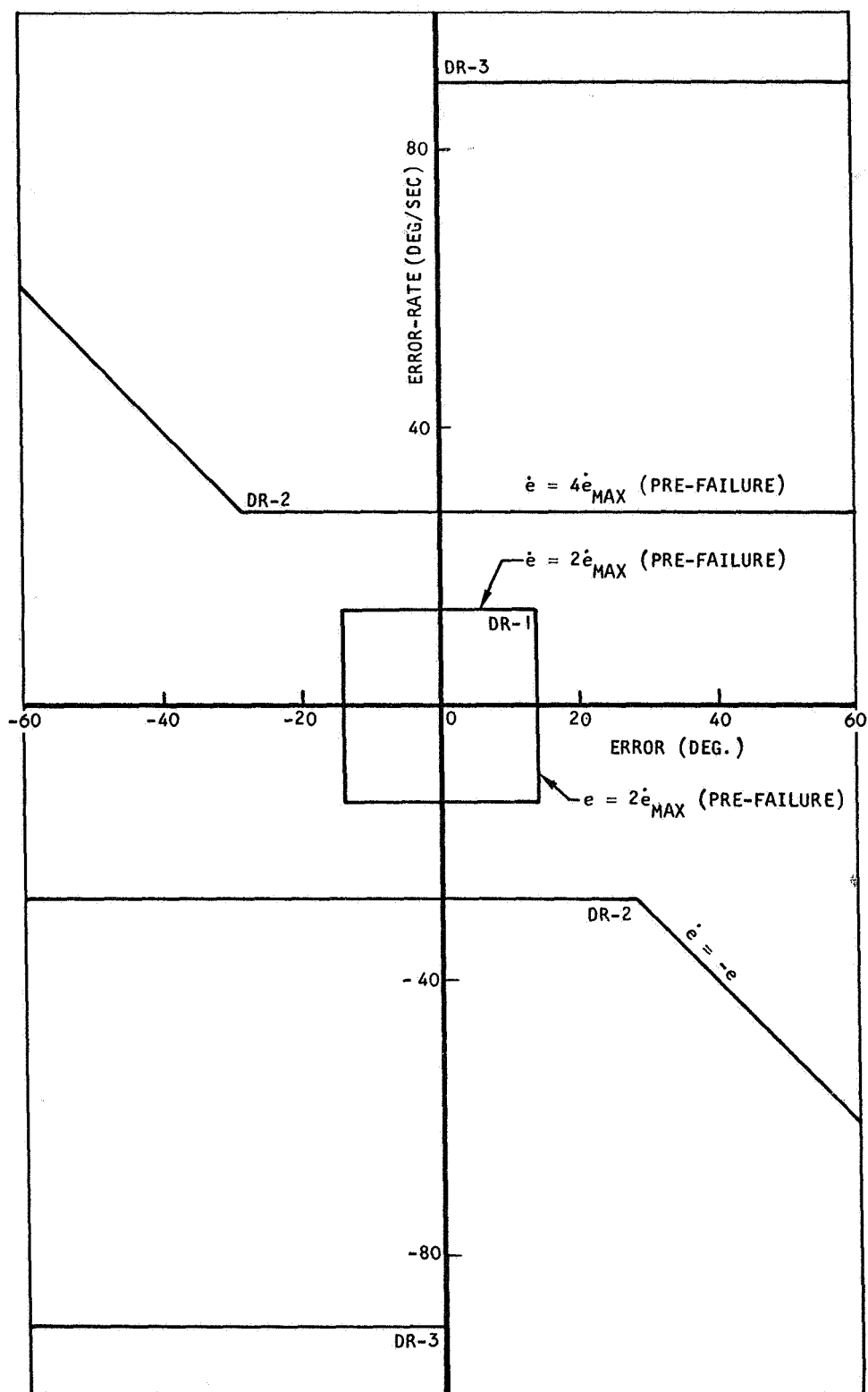


Figure 5. —Decision regions for detection and pattern recognition in the supervisory control algorithm.

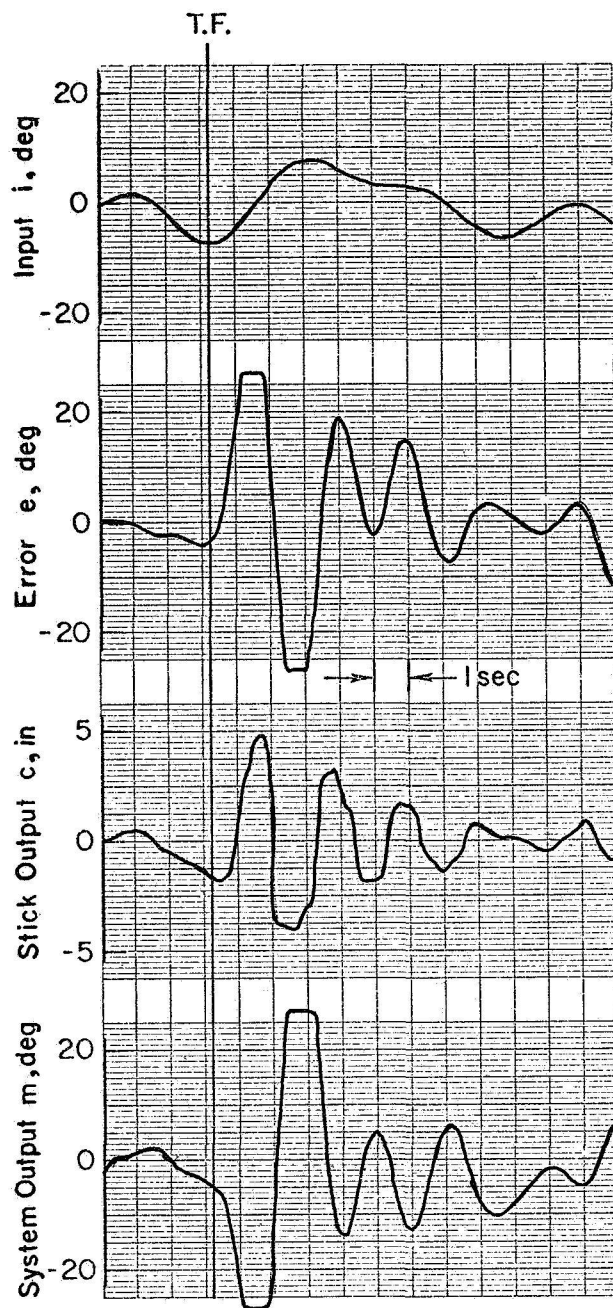


Figure 6. —Human operator response—soft failure—level A to C.

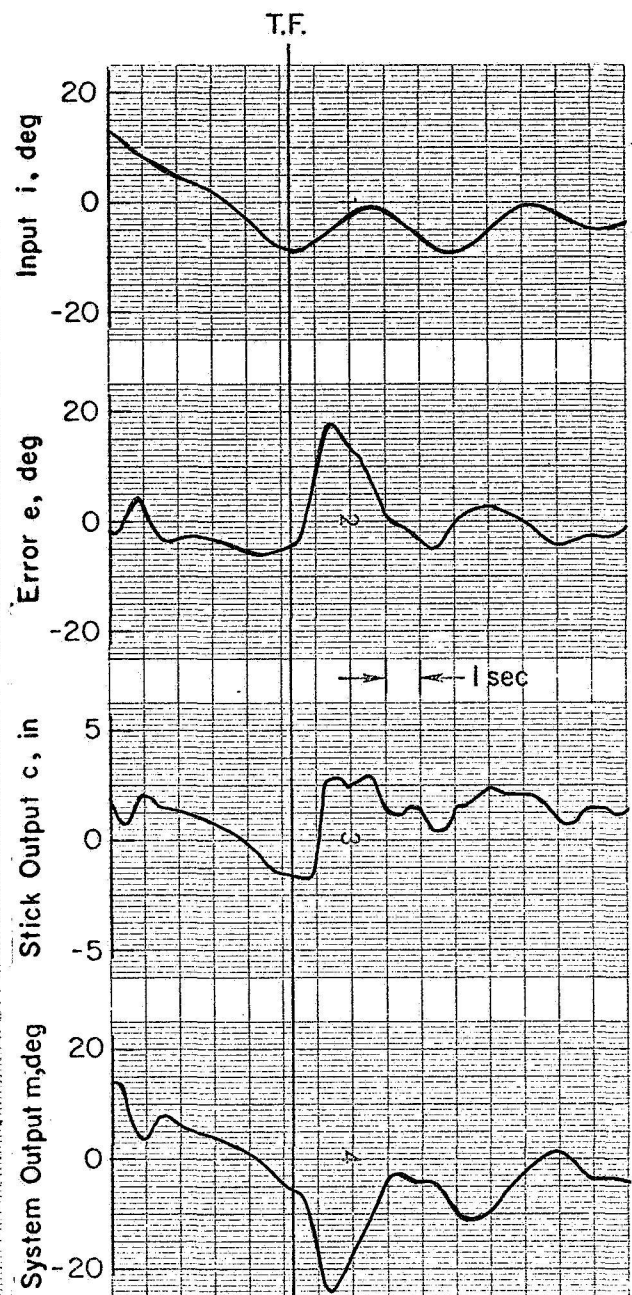


Figure 7. —Human operator response—hard failure—level A to B.

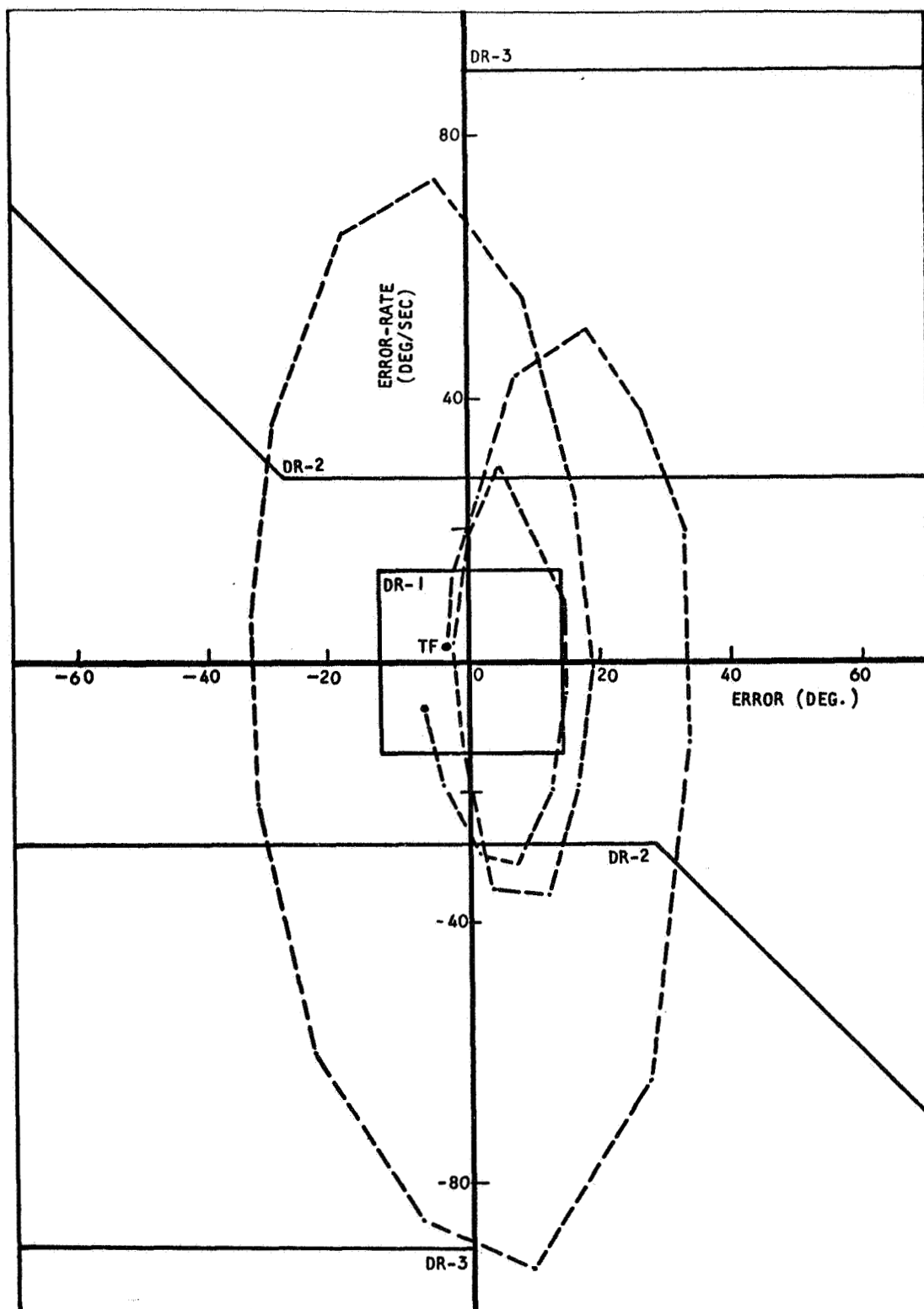


Figure 8.—Phase-plane plot of failure response shown in figure 6.

bounds then occurrence of a failure is assumed. This region is defined as decision region DR-1 and is shown in figure 5. The proposed detection criterion is deterministic to a first approximation and can easily be made stochastic by introducing appropriate randomness in the detection scale factor (say $(2 \pm \text{noise}) \times \text{stored upper bounds}$).

Once a failure is detected the operator model is switched to the structure appropriate to augmentation level B (to augmentation level C for set 2 experimental situation as shown in fig. 4). There are three likely postfailure dynamics corresponding to levels B, C, or D.

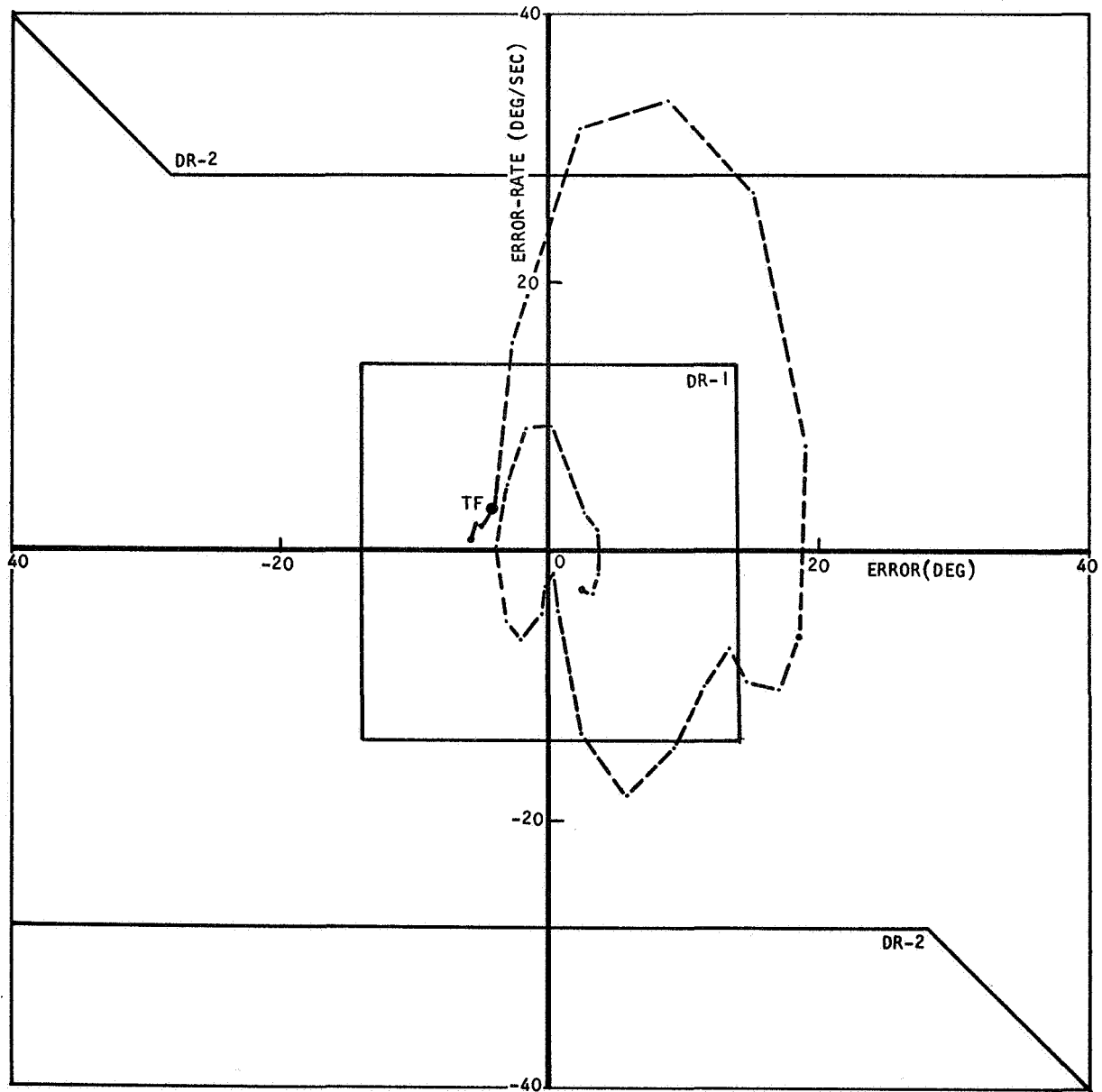


Figure 9. —Phase-plane plot of failure response shown in figure 7.

If the postfailure augmentation is level B, then the operator model is appropriate to it and the phase trajectory moves toward the origin, after the error rate changes sign. This is apparent in human-operator data of figures 7 and 9. Notice that the trajectory following the reversal in error rate never enters the decision region DR-2. In case of a hard failure as in figures 7 and 9, the operator stick response approaches a steady bias of ± 1.5 inches (step transient at failure is ± 1.5 inches in height) while the error and error rate become small. The model as presented in figure 3 will give the desired stick bias, but has a steady-state error equal to the reciprocal of the low frequency human operator model gain for a unit step failure transient. However, this value is small ($1.5 \times 57.3 / 17.2 = 5$ degrees for $Y_{H.O}(j\omega) = 17.2 e^{-.24(j\omega)}$) enough to go unnoticed for a few seconds following failure. This bias in the error and stick magnitudes gives the necessary cue for introducing a trim (dc bias) in the stick, thereby reducing the error to values near zero. Hereafter follows adjustment of model parameters to desired steady-state values.

If after detection, the postfailure augmentation level is not B, it must be either C or D. The number of alternatives has now been reduced from three to two. The overall system is still unstable in an oscillatory divergent mode. This causes the error rate to increase rapidly in magnitude after reversing its sign and the phase trajectory enters decision region DR-2. This is seen in human-operator data for a soft failure in augmentation from level A to level C as seen in the example phase-plane data of figure 8. This is in contrast to figure 9 where the trajectory never entered DR-2. It is postulated here that if the trajectory, after detection and error-rate reversal, does enter DR-2, then the postfailure augmentation level cannot be B and must therefore be C or D. The operator-model structure is next switched to one appropriate to level C. If it is C, the error and error rate go to zero and the model parameters are adjusted to desired steady-state values.

If the postfailure dynamics correspond to level D, then the overall system being still unstable causes the trajectory to enter DR-4, following error-rate reversal, whereby the identification of augmentation level as D is concluded. The operator-model structure is appropriate to level D, but the time delay must be reduced. In the human operator this amounts to reducing the time delay to the physiological limit, thereby causing a loss of control in one out of three trials for such failures.

Having provided a framework for the supervisory-control algorithm, results of failure response of the human-operator adaptive model of figures 2 and 3 are presented next.

DIGITAL SIMULATION OF HUMAN-OPERATOR ADAPTIVE MODEL (INCORPORATING THE SUPERVISORY CONTROL LOGIC) TO FAILURES

The complete adaptive model of the human operator was simulated on a digital computer using the MIMIC program for solving systems of differential equations. Results for two sample cases corresponding to those in figures 6 and 7 are presented in figures 10 and 12, respectively. Note, however, that the model input was generated digitally and is not a duplicate in shape of either of the data inputs in figures 6 or 7. The corresponding phase-plane plots for the model response are shown in figures 11 and 13, respectively. The decision regions DR-1, 2 and 3 used in the simulations are shown superimposed on the phase-plane plots of figures 11 and 13. For convenience, let

T1=Time at which the operator-model structure switches to that appropriate to augmentation level B

T2=Time at which the operator-model structure switches to that appropriate to augmentation level C

T_3 = Time at which the operator-model structure switches to that appropriate to augmentation level D

Depending on the type of failure, T_2 and/or T_3 may be infinite, meaning thereby that the phase trajectory never enters decision regions DR-2 and/or DR-3. The location of these instants on the phase trajectory is identified in figures 11 and 13, along with the human-operator models (including parameter values chosen) for the successive time intervals. The model parameters chosen for these transitions correspond to values appropriate to the respective steady-state conditions. This is not essential, and one can generate a family of different transition responses for the same input conditions by varying the parameters chosen in successive model structures and also by adjusting the quantitative features (not shape) of the three decision regions.

The model stick response of figure 10 shows bang-bang features with the first bang lagging the first error peak which is typical of human-operator response to such failures. Similarly, the model response to a hard failure as in figure 12, gives an average stick bias and small errors following completion of adaptation that corresponds to human-operator data for this type of failure. Other measures of the transition response such as retention time, peak error, maximum error rate, and setting time also fall in the magnitude range observed in operator data.

These similarities in model response to human-operator data demonstrate the feasibility of the proposed human-operator adaptive model. As a final example, a comparison of model response to human-operator response for identical data inputs is presented in figures 14 and 15. Figure 14 shows the time responses while figure 15 gives the corresponding phase-plane portraits. The human-operator models used for the successive structures and the decision-region boundaries used for pattern recognition are given in figure 15. These model parameters are by no means unique, and computer simulations with different sets of parameters

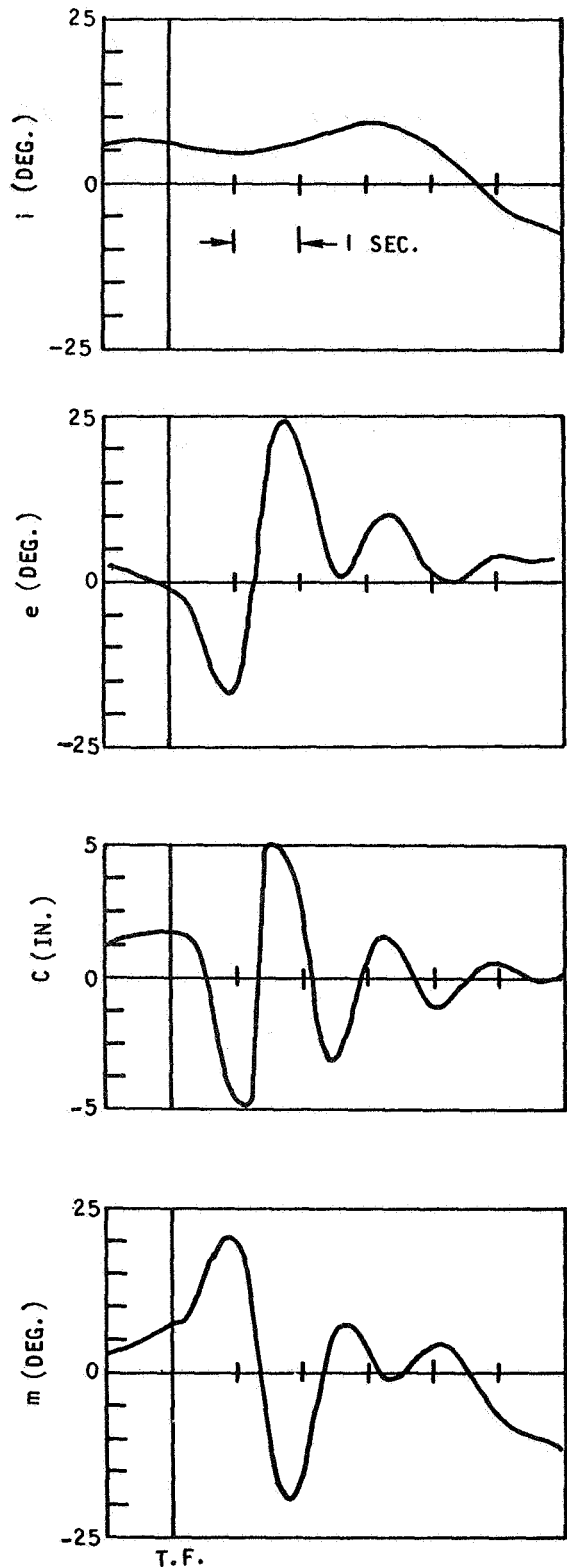


Figure 10. — Adaptive model response to a soft failure from augmentation level A to C.

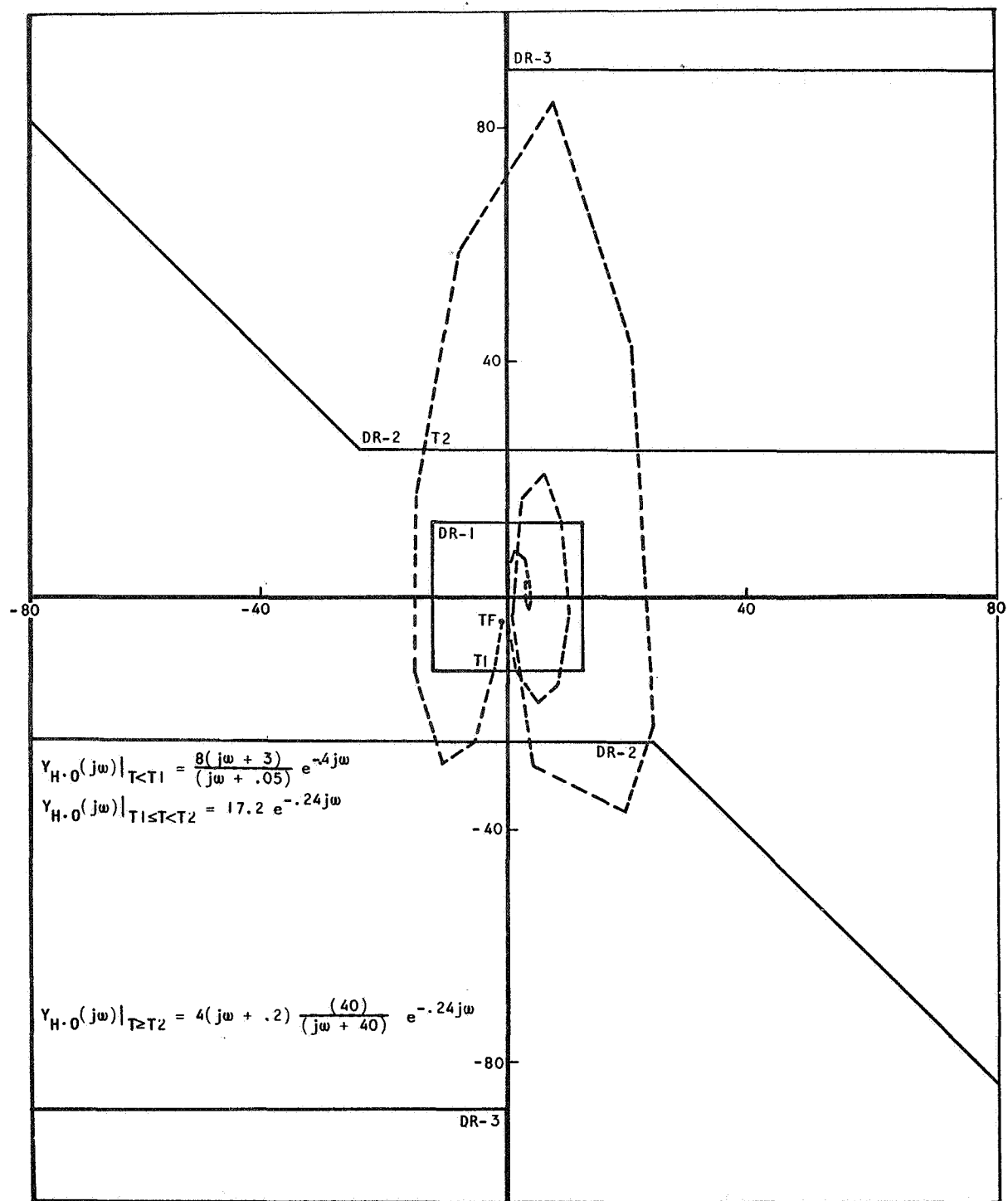


Figure 11. —Phase-plane plot of model failure response shown in figure 10.

(not given in this paper) indicate that variations in the time traces may be obtained while keeping the same general characteristics of the response. Notice that the model response matches operator response very well until $T=TF+3$ seconds, after which the operator response seems delayed with respect to the model response. One reason for the observed discrepancy beyond these seconds after failure may be that the human operator changed his parameters—specifically the time delay and lead time constant due to some disturbance or that he made a mistake in estimating system error and error rate. The model response is faster and has smaller steady-state errors than the human-operator response. Thus the human-operator response may be considered to be suboptimal relative to the proposed adaptive model response. However, it must be emphasized that the human-operator response to failures discussed herein is stochastic and that any model-response comparisons with actual human-operator data must be done in a statistical sense. It is appropriate to say that the model presented herein has the capability to demonstrate such statistical equivalence, given adequate data.

CONCLUSIONS

The human-operator adaptive model of figures 2 and 3 (or 4) is an example of an adaptive controller in which the time constant of the adaptive or higher order loop T_a is of the same order of magnitude as the overall system time constant T_s . Adaptive systems presented so far assume a wide range of magnitude between these two time constants (ref. 15). For $T_s < T_a$, the system may be analyzed by using average input and plant variations. Peak seeking and estimating adjustment mechanizations belong to this type. If $T_s > T_a$, it is assumed that the plant parameter variations are known, and model reference systems have been used for this purpose. No such assumptions can be made when $T_s \doteq T_a$ and hence this problem has proved quite difficult. Feldbaum's dual-control theory (ref. 16) has been applied to

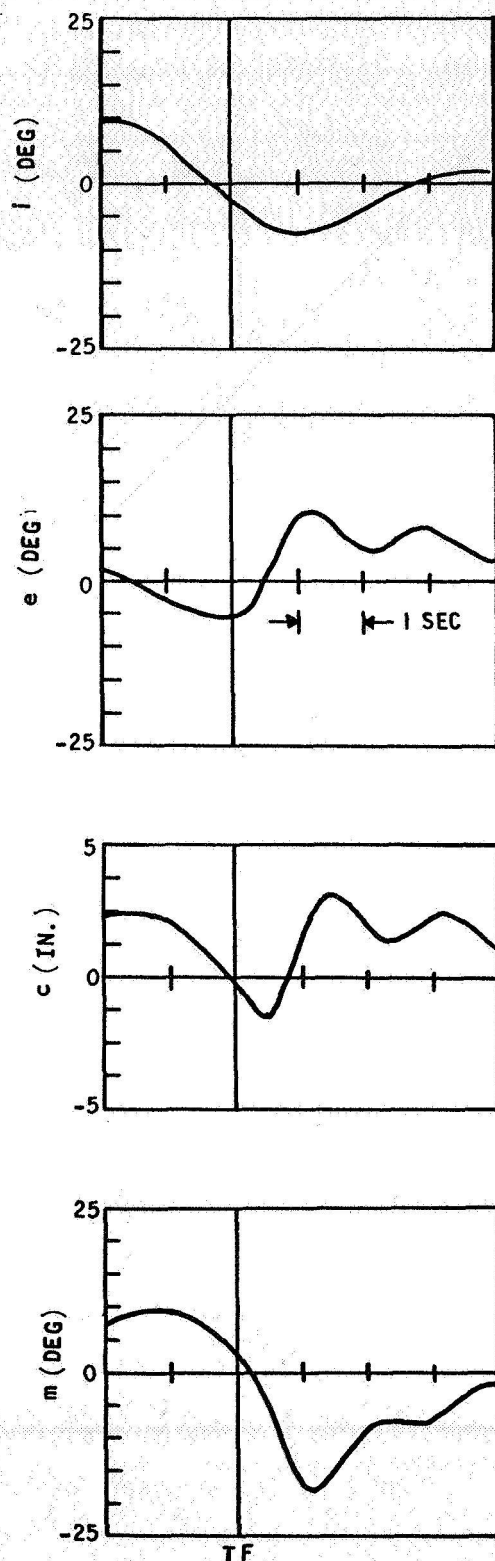


Figure 12. — Adaptive model response to a hard failure from augmentation level A to B.

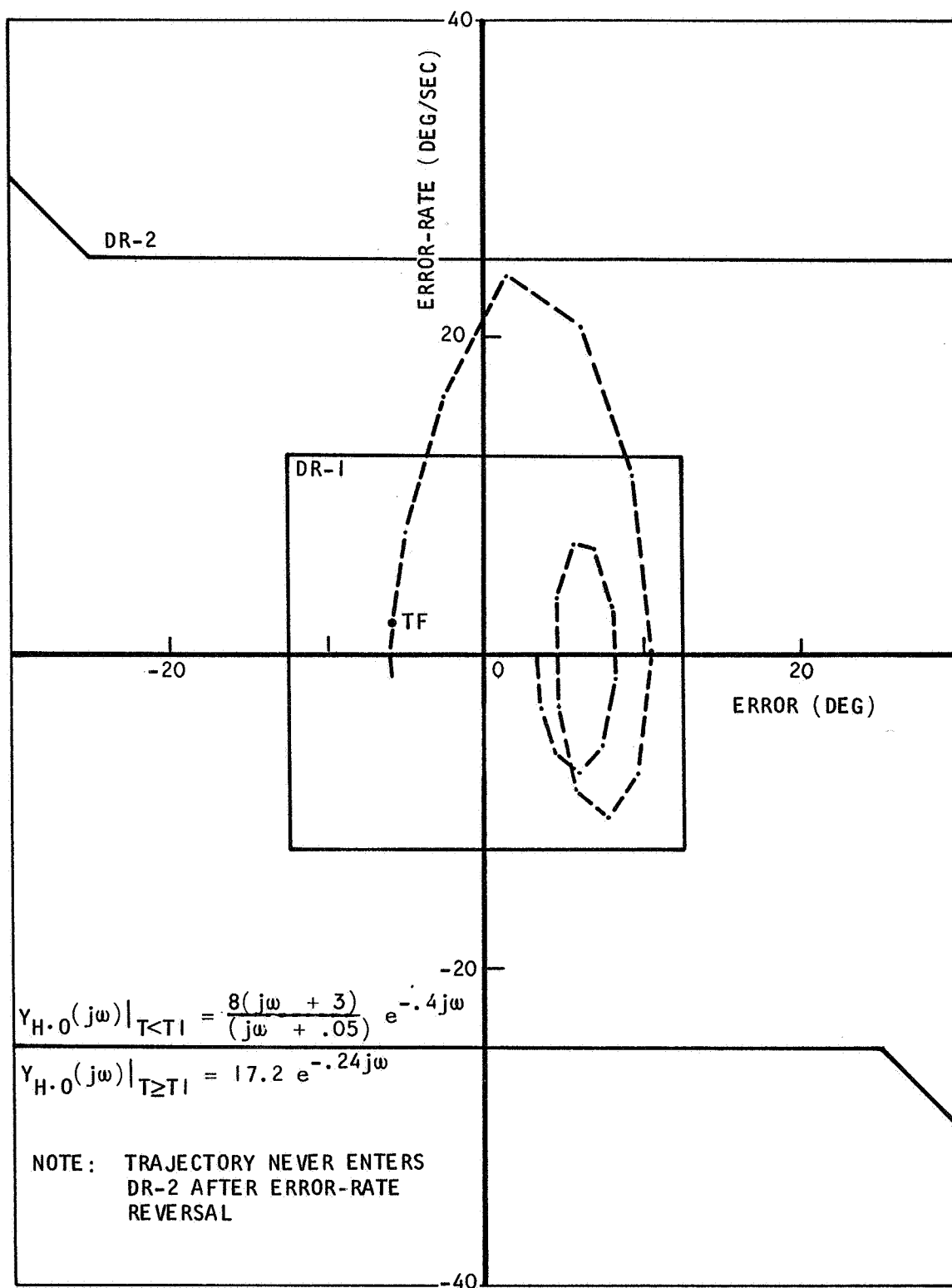


Figure 13. —Phase-planeplot of model failure response shown in figure 12.

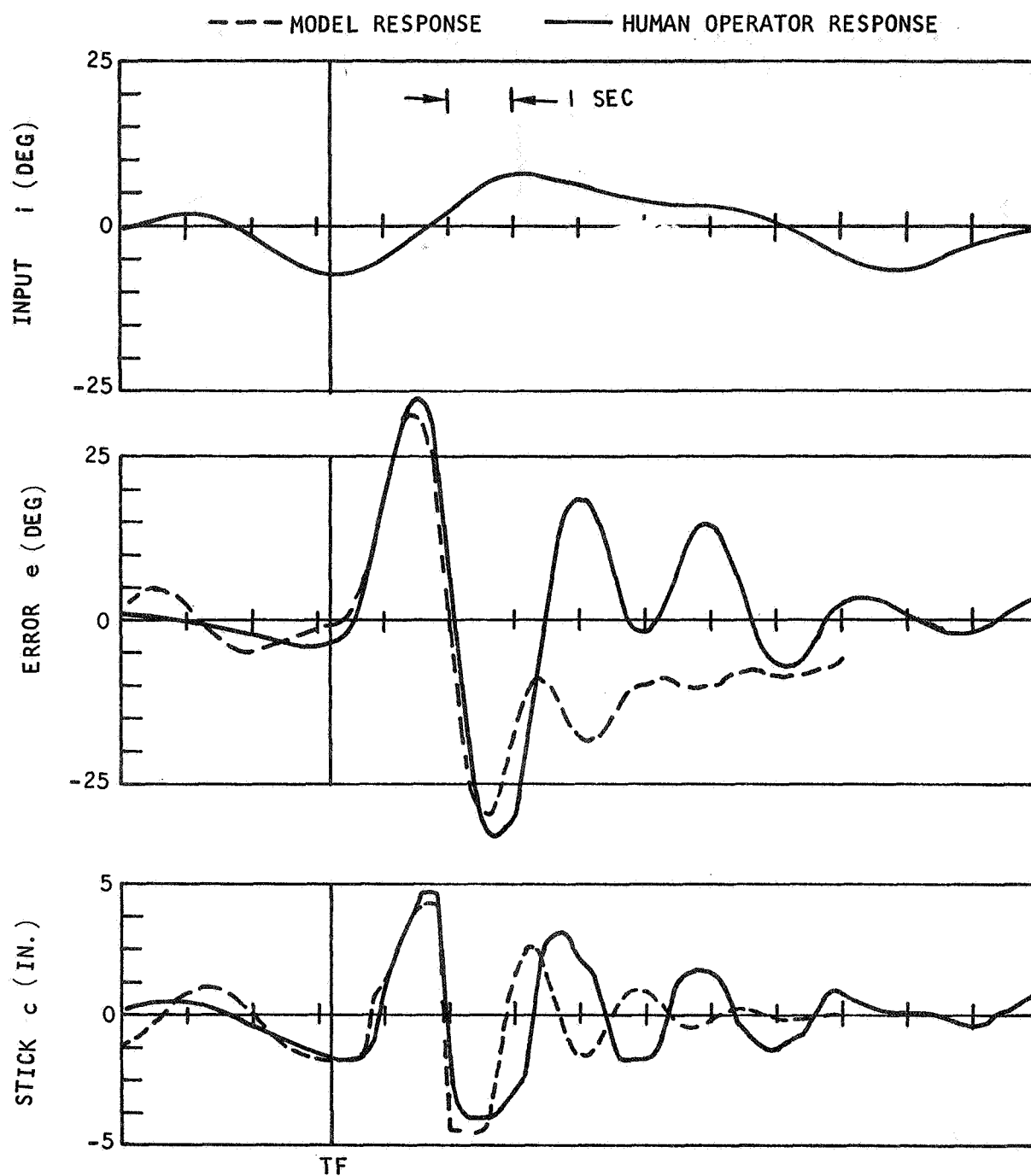


Figure 14. — Comparison of adaptive response with human operator response for a soft failure from augmentation level A to level C with the same data input in both cases.
 (Note: human-operator response data is the same as in fig. 6.)

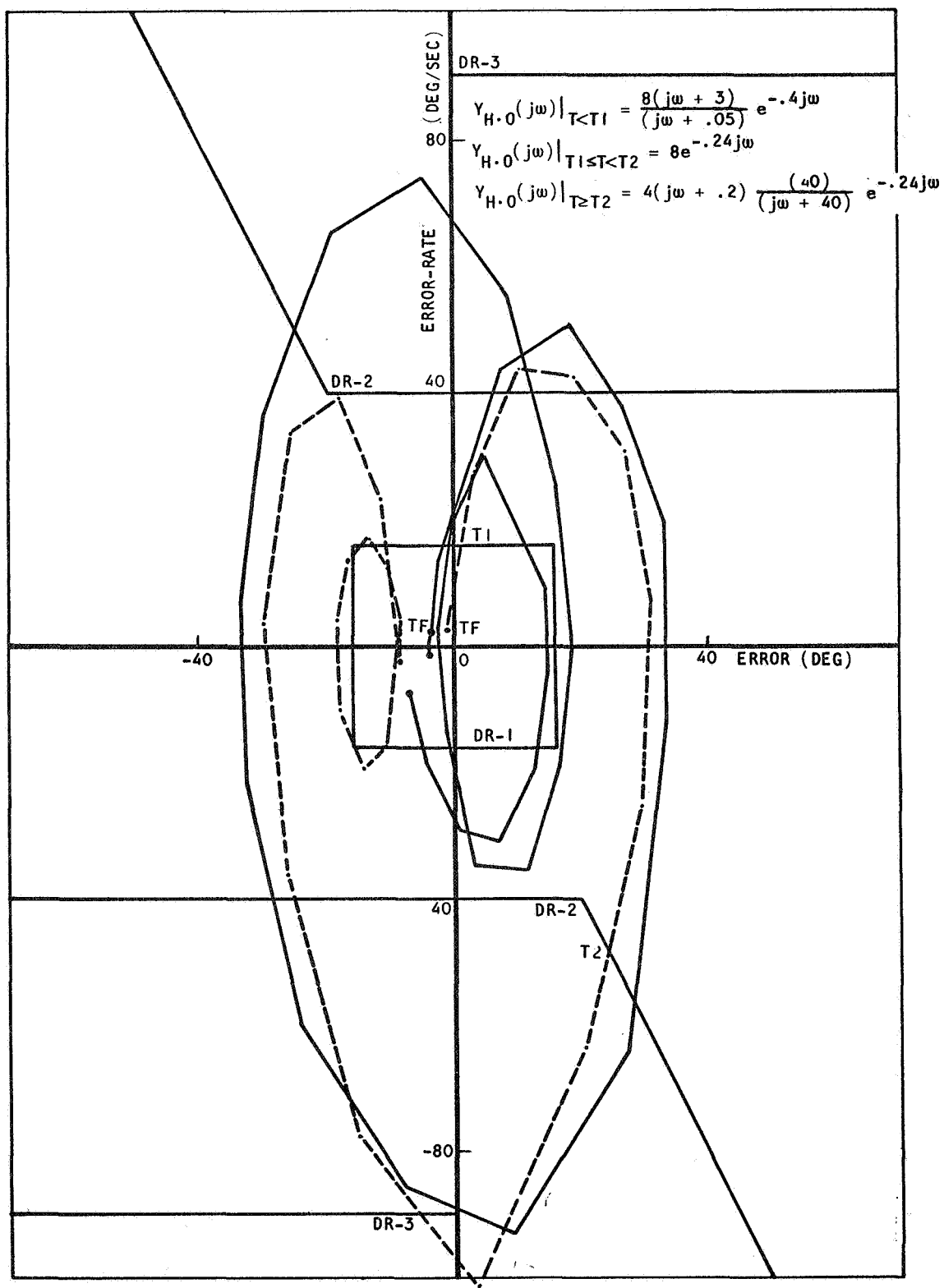


Figure 15. —Phase-plane plot of failure responses shown in figure 14.

simple problems of this sort, but has failed to yield solutions to any meaningful applications. Recent attempts to apply sensitivity methods without using a plant model seem encouraging (ref. 17). However, results so far are rather limited in scope. In short, the model proposed herein attempts to solve the problem discussed above (when $T_s = T_a$) using the concepts of pattern recognition and decision processes. The human-operator model is stochastic in nature by virtue of the run-to-run variability introduced in the decision region boundaries, the parameters of the various successive model structures, and, finally, the measurement noise in the observed variables' error and error rate.

The same general principles may be applied to any other particular time-varying control situation since there must be some inherent ordering in the given nonstationary situation. This ordering can be identified using pattern recognition as evidence and the operator structure then modified sequentially to achieve the desired response. The adaptation strategy being modeled is assumed to have reached steady state via learning.

It is recommended that learning itself be modeled using some kind of penalty-reward approach to arrive at the steady-state, quasi-deterministic, adaptive-control strategy, such as the one suggested in this paper. Research results of reference 18 may be worth applying to this problem.

REFERENCES

1. Sheridan, T. B.: Time-Variable Dynamics of Human Operator Systems. AFCRC-TN-60-169 (ASTIA AD-237045), Mar. 1960.
2. Sheridan, T. B.: Studies of Adaptive Characteristics of the Human Controller. ESD-TDR-62-351, Dec. 1962.
3. Young, L. R.; Green, D. M.; Elkind, J. I.; and Kelly, J. A.: The Adaptive Dynamic Response Characteristics of the Human Operator in Simple Manual Control. NASA TN D-2255, 1964.
4. Elkind, J. I.; Kelly, J. A.; and Payne, R. A.: Adaptive Characteristics of the Human Controller in Systems Having Complex Dynamics. IEEE Proc. Fifth Natl. Symp. Human Factors Electronics, May 1964.
5. Elkind, J. I.; and Miller, D. C.: On the Process of Adaptation by the Human Controller. IFAC Meeting, London, England, 1966.
6. Elkind, J. I.; and Miller, D. C.: Adaptive Characteristics of the Human Controller of Dynamic Systems. AFFDL-TR-66-60, July 1966.
7. Sadoff, M.: A Study of a Pilot's Ability to Control During Simulated Stability Augmentation System Failures. NASA TN D-1552, 1962.
8. Miller, D. C.: A Model for the Adaptive Response of the Human Controller to Sudden Changes in Controlled Process Dynamics. B.S. and M.S. Thesis, MIT, 1965.
9. Weir, D. H.; and Phatak, A. V.: Model of Human Operator Response to Step Transitions in Controlled Element Dynamics. NASA CR-671, Jan. 1967.
10. Hess, R. A.: The Human Operator as an Element in a Control System With Time Varying Dynamics. AFFDL-FDCC-TM-65-34, June 1965.
11. Hess, R. A.: An Investigation of the Human Operator as an Element in Both Time Variant and Equivalent Time Invariant Systems. AFFDL-FDCC-TM-65-42, Sept. 1965.
12. Weir, D. H.; and Johnson, W. A.: Pilot Dynamic Response to Sudden Flight Control System Failures and Implication for Design. Systems Tech., Inc. TR-165-1, Oct. 1967.

13. McRuer, D.; Graham, D.; Krendel, E.; and Reisener, W., Jr.: Human Pilot Dynamics in Compensatory Systems: Theory, Models, and Experiments with Controlled Element and Forcing Function Variations. AFFDL-TR-65-15, July 1965.
14. Bellman, R.: Adaptive Control Processes—A Guided Tour. Princeton Univ. Press, 1961.
15. Chang, S. S. L.: On the Relative Time of Adaptive Processes. IEEE Trans. on Auto-Control, vol. AC-10, no. 1, Jan. 1965.
16. Feldbaum, A. A.: Dual Control Theory Problems. Proc. Second Intern. Cong. on Auto-Control, Basel, Switzerland, Sept. 1963.
17. Kokotovic, P.; et al.: Sensitivity Method in the Experimental Design of Adaptive Control Systems, Proc. Third Intern. Cong. on Auto-Control, London, June 1966.
18. Preyss, A. E.; and Meiry, J. L.: Stochastic Modeling of Human Learning Behavior. Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967, pp. 327-349.

21. On the Dynamic Response of the Human Operator to Transient Inputs*

*Anil V. Phatak and David H. Weir
Systems Technology, Inc*

A dual-mode controller is used to describe the operator's dynamic response to inputs containing both deterministic (transient) and random-appearing signals. It includes the usual quasi-linear describing function plus a feedforward path which produces the transient response. New data are presented for step responses with effective controlled elements of K , K/s , K/s^2 , and K/s^3 . They indicate that previously hypothesized sample-data models are primarily valid for pure gain controlled elements. The data show that step responses for a trained operator have portions which are nearly time optimal, and that this is one possible idealization for the feedforward, transient-response path in the model.

INTRODUCTION

The simplest nonstationary control situation is one in which a highly trained, nonalerted subject operating a constant-coefficient linear-controlled element in a compensatory system is confronted with one randomly occurring step input. In spite of the simple circumstances, the overall behavior is complicated when a variety of controlled elements are considered. The skilled operator's output is peculiar to each controlled-element structure, although the system response in all cases tends to follow the forcing function after a time delay. The system output to a step-input forcing function of amplitude A sketched in figure 1 is typical. Extant data indicate that the response is not that which obtains when a step is applied to the quasi-linear describing-function model derived from stationary situations with random-appearing inputs. The settling time increases monotonically with increase in the order of the controlled element, and the first overshoot is flat-topped, indicating the presence of nonlinearity in the loop.

Thus, for tracking random inputs with occasional step inputs, a dual-mode model of the operator is appropriate as shown in reference 1. The basic structure of the model is given in figure 2. The quasi-linear steady-state path is one used for tracking random inputs, while the feedforward element operates on the step. The basic structure also incorporates

*This research was accomplished in part under contract AF 33(615)-3652 for the Air Force Flight Dynamics Laboratory.

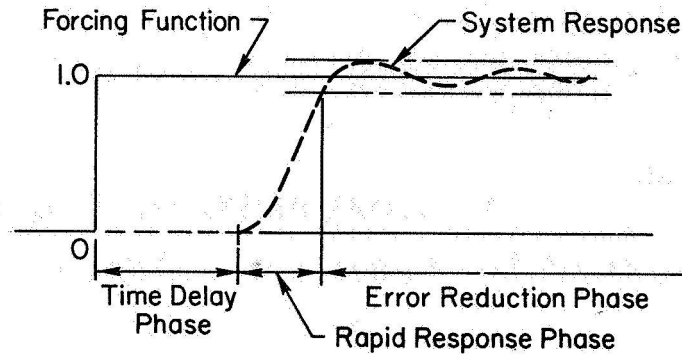


Figure 1. — Typical system step response.

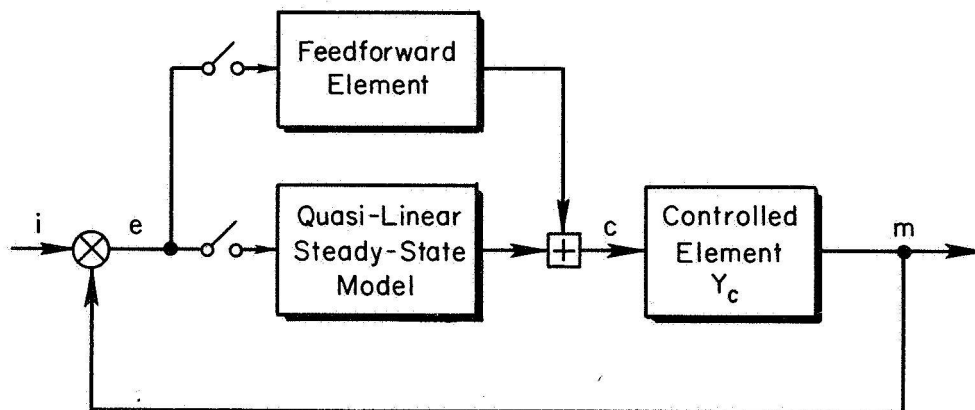


Figure 2. — Structure of the dual-mode model.

mode switches for the two pathways. In terms of the temporal phases of figure 1, the successive action structures of the dual-mode model are

- (1) Transition from quasi-linear mode to feedforward mode, corresponding to the time-delay phase
- (2) Patterned feedforward response, corresponding to the rapid-response phase
- (3) Quasi-linear mode, corresponding to the error reduction phase

The duration of the time-delay phase has some minimum value, and its unimodal distribution is skewed to the right. The time delay τ_s is generally longer than the steady-state effective time delay τ_e in tracking. In particular

$$\tau_s = \tau_e + \tau_d \quad (1)$$


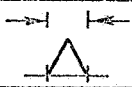


where τ_s is the time-delay phase duration, τ_e is the effective time delay in steady-state tracking, and τ_d is the decision time. During the decision time, the operator makes the pertinent decisions regarding the shape and magnitudes of the feedforward response. At the end of the time-delay phase, the feedforward element generates the proper response to the step-input command, giving the rapid-response phase. At the end of this period, the error is small, and the operator switches to a quasi-linear tracking mode in the error-reduction phase.

The quasi-linear controller is well known, and the following is confined to the parallel feedforward path utilized in response to step inputs. Prior models have used sampled-data elements in the feedforward, and these are reviewed briefly in the next section. Then step response data are presented. Finally, a time-optimal form is described which better describes the data.

SAMPLED-DATA MODELS FOR STEP RESPONSE

Past models for the operator's step-response behavior have used sampled-data elements for the feedforward. A recent version, due to Elkind (ref. 2), is generally a hybrid of the two-channel pursuit/saccadic control logic of the eye-movement model of reference 3 and the velocity-triangle force program of a hand-movement model from reference 4. The model incorporates a dual-rate sampler that provides a high sampling rate while waiting then slows down after the input is sensed to permit the system to complete its response before the next sample is taken. The model is extended in reference 1 to account for higher order controlled elements. The salient features of this sampled-data model are summarized in table 1. The delay-time distribution of the model is rectangular and uniformly distributed between 0.15 and 0.30 second, in contrast with past data (e. g., ref. 5) which indicate a near-gaussian distribution of the inverse time delay. This was the major piece of preexperimental evidence that the model did not explain. The time T_c to complete the force program is constant at one sampling interval (0.15 second.)

TABLE 1.—SUMMARY OF SAMPLED-DATA MODEL PREDICTIONS

CONTROLLED ELEMENT Y_c	TIME TO COMPLETE FORCE PROGRAM (AT STICK)	WAVEFORM OF OPERATOR RESPONSE (STICK DEFLECTION)
K_c	0.15 sec	 : Second-order response to step
K_c/s	0.15 sec	 : Triangle
K_c/s^2	0.15 sec	 : Double pulse
K_c/s^3	0.15 sec	 : Impulse train

EXPERIMENTAL DATA

Step-response data for various controlled elements and differing magnitude step inputs were obtained in order to confirm or modify the sampled-data model which existed at the outset. Typical examples of step-response data for $Y_c = K_c$, K_c/s , K_c/s^2 , and K_c/s^3 are presented in figure 3(a), (b), (c), and (d), respectively. For each controlled element the value of the gain K_c was picked on the basis of best subjective opinion rating for the overall system.

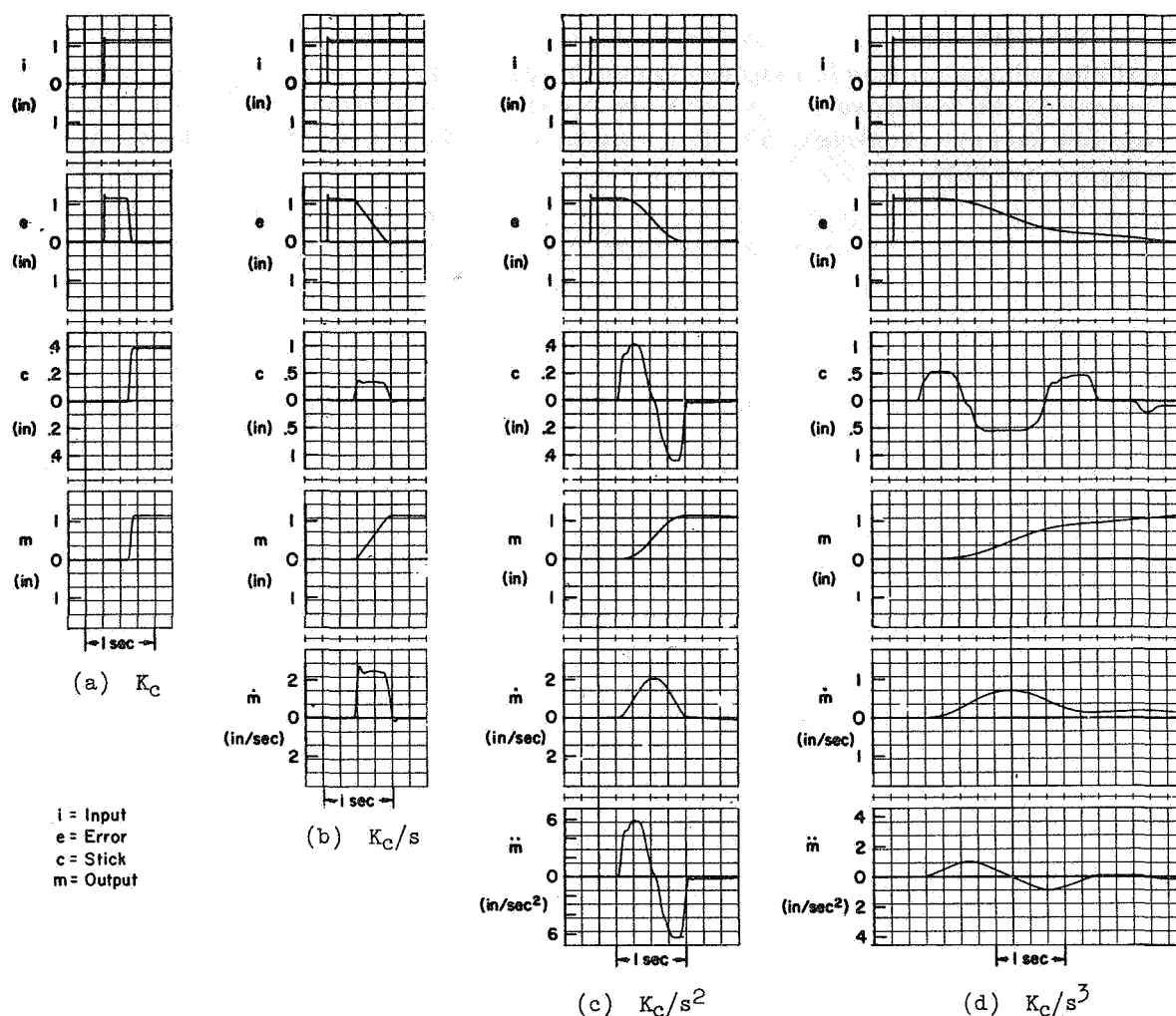


Figure 3.—Typical step response data (a) K_C , (b) K_C/s , (c) K_C/s^2 , (d) K_C/s^3 .

The important step-response parameters are obtained by averaging data extracted from several¹ runs, and the means and standard deviations are given in table 2. Some differences are evident between the sampled-data model predictions (table 1) and the data. The model predicted a uniform distribution of delay time between 0.15 and 0.3 second for all Y_C , while the data show different delay times (with low variance) for each Y_C . The model predicted a constant time (0.15 second) to complete the force program, and the data show the time increasing monotonically with controlled element order. A triangular stick pulse was predicted for K_C/s , while the data show a more rectangular pulse. The model gave three impulses (plus, minus, plus) for K_C/s^3 , but the data shows three approximately rectangular pulses of about equal magnitude, and the middle pulse has about twice the duration of the first and third. The principal differences are in the time to complete the force program and the fact that the data show, for all controlled elements,

¹Each entry in table 2 is the result of approximately 15 experimental runs with a highly trained and motivated operator.

TABLE 2. — EXPERIMENTAL STEP RESPONSE PARAMETERS

[Magnitude in inches, time in seconds. Mean at top, standard deviation at bottom.]

CONTROLLED ELEMENT y_c	INPUT HEIGHT	SYSTEM RISE TIME	DELAY TIME	FIRST PULSE $c(t)$		SECOND PULSE $c(t)$		THIRD PULSE $c(t)$		TIME TO COMPLETE FORCE PROGRAM	COMMENTS
				Height	Width	Height	Width	Height	Width		
2.86	$\frac{8}{7}$	0.072 0.013	0.347 0.035	0.360 0.014	—	—	—	—	—	0.134 0.020	Step deflec- tion of stick
2.86	$\frac{10}{7}$	0.055 0.010	0.310 0.041	0.405 0.041	—	—	—	—	—	0.132 0.015	
$\frac{7.1}{s}$	$\frac{6}{7}$	0.420 0.063	0.320 0.032	0.250 0.036	0.518 0.061	—	—	—	—	0.518 0.061	Single-pulse deflection of stick
$\frac{7.1}{s}$	$\frac{8}{7}$	0.432 0.048	0.352 0.029	0.283 0.031	0.523 0.045	—	—	—	—	0.523 0.045	
$\frac{7.1}{s}$	$\frac{10}{7}$	0.545 0.148	0.355 0.073	0.303 0.046	0.590 0.060	—	—	—	—	0.590 0.060	Double-pulse deflection of stick (\pm). Approximately equal ampli- tude bangs and equal pulse width.
$\frac{14.3}{s^2}$	$\frac{6}{7}$	0.905 0.100	0.347 0.063	0.201 0.032	0.513 0.059	0.157 0.034	0.640 0.069	—	—	1.153 0.059	
$\frac{14.3}{s^2}$	$\frac{8}{7}$	0.839 0.063	0.367 0.063	0.297 0.029	0.524 0.046	0.267 0.052	0.540 0.068	—	—	1.063 0.060	Triple-pulse deflection of stick
$\frac{14.3}{s^2}$	$\frac{10}{7}$	0.675 0.065	0.323 0.057	0.452 0.122	0.510 0.054	0.422 0.125	0.561 0.107	—	—	1.071 0.155	
$\frac{3.6}{s^2}$	$\frac{8}{7}$	2.542 0.448	0.378 0.045	0.420 0.094	0.755 0.112	0.443 0.150	1.212 0.216	0.320 0.139	0.825 0.195	2.792 0.330	
$\frac{3.6}{s^2}$	$\frac{10}{7}$	2.145 0.868	0.403 0.076	0.620 0.116	0.710 0.053	0.753 0.166	1.132 0.140	0.517 0.166	0.965 0.386	2.807 0.385	

stick motions which are more nearly rectangular than the triangular or impulsive predictions of the odd-order controlled elements.

TIME OPTIMAL MODEL FOR STEP RESPONSE

Differences between the sampled-data model and the data spur a search for new model forms which will account for the experimental results. The approximately rectangular or bang-bang shape of the stick movements (operator response) suggests an optimality of the operator and his related performance indices.

An attractive explanation is that the operator is optimal or suboptimal relative to the minimum time criterion. Consider an n th order single-input/single-output control system with $|c(t)| \leq M$, where the scalar M may represent either a physical limit on the stick deflection or, more likely, an implicit magnitude restraint imposed by the operator for the given situation. For $i(t) = \text{constant}$, the time-optimal control has the following properties:

The control $c(t)$ is bang-bang, i. e., $c(t) = +M$ or $-M$.

There are at most $(n-1)$ switchings (i. e., $+M$ to $-M$, or vice versa) for systems with n real eigenvalues.

The switching logic is dependent on the order of the controlled element. In general, the switching surface is a nonlinear function of the state variables.

For a given initial condition of the state variables, there is one unique control $c(t)$.

Then for the problem at hand, there is a specific type of initial condition of the system-state vector, namely

$$\begin{aligned} e(0) &= \text{input height, } A \\ \dot{e}(0) &= 0 \\ \ddot{e}(0) &= 0 \\ &\vdots \end{aligned}$$

The terminal state is the origin.

In order to measure the degree to which the available step-response data is time-optimal, invariance conditions can be obtained for each controlled element by solving a two-point boundary-value problem. These are stated in table 3, omitting their derivation (ref. 6). The symbol $(\cdot)_0$ denotes the time-optimal value of the parameter in parentheses.

Assuming a wide-band neuromuscular-system response, the ideal time-optimal step-response characters for $Y_c = K_c$, K_c/s , K_c/s^2 , K_c/s^3 are shown in figure 4(a), (b), (c), and (d), respectively. Note that the smoothing effect of the neuromuscular system would round off the corners in the figure 4 responses yielding results similar to that in figure 3. Further, the control movement starts after the end of the delay time phase, and time optimality pertains to that period of control only.

The invariance conditions for the time-optimal model interpretations are compared with the experimental results in table 4. The values of the invariance condition computed from the experimental results are denoted by $(\cdot)_e$. The last column in table 4 gives the percentage error from time-optimal. The comparison between data and model is seen to be quite good, and it appears reasonable to conclude that the step-response behavior of highly trained, attentive operators is nearly time-optimal. Variations in the switching logic appearing in actual data do not affect the terminal zero-error condition substantially, but

TABLE 3.—INVARIANCE CONDITIONS FOR TIME OPTIMALITY

CONTROLLED ELEMENT, Y_c	INVARIANCE CONDITION FOR TIME OPTIMALITY
K_c	$= A/K_c$
K_c/s	$(T_c M)_O = A/K_c$
K_c/s^2	$(T_c M^{1/2})_O = 2(A/K_c)^{1/2}$
K_c/s^3	$(T_c M^{1/3})_O = (32A/K_c)^{1/3}$

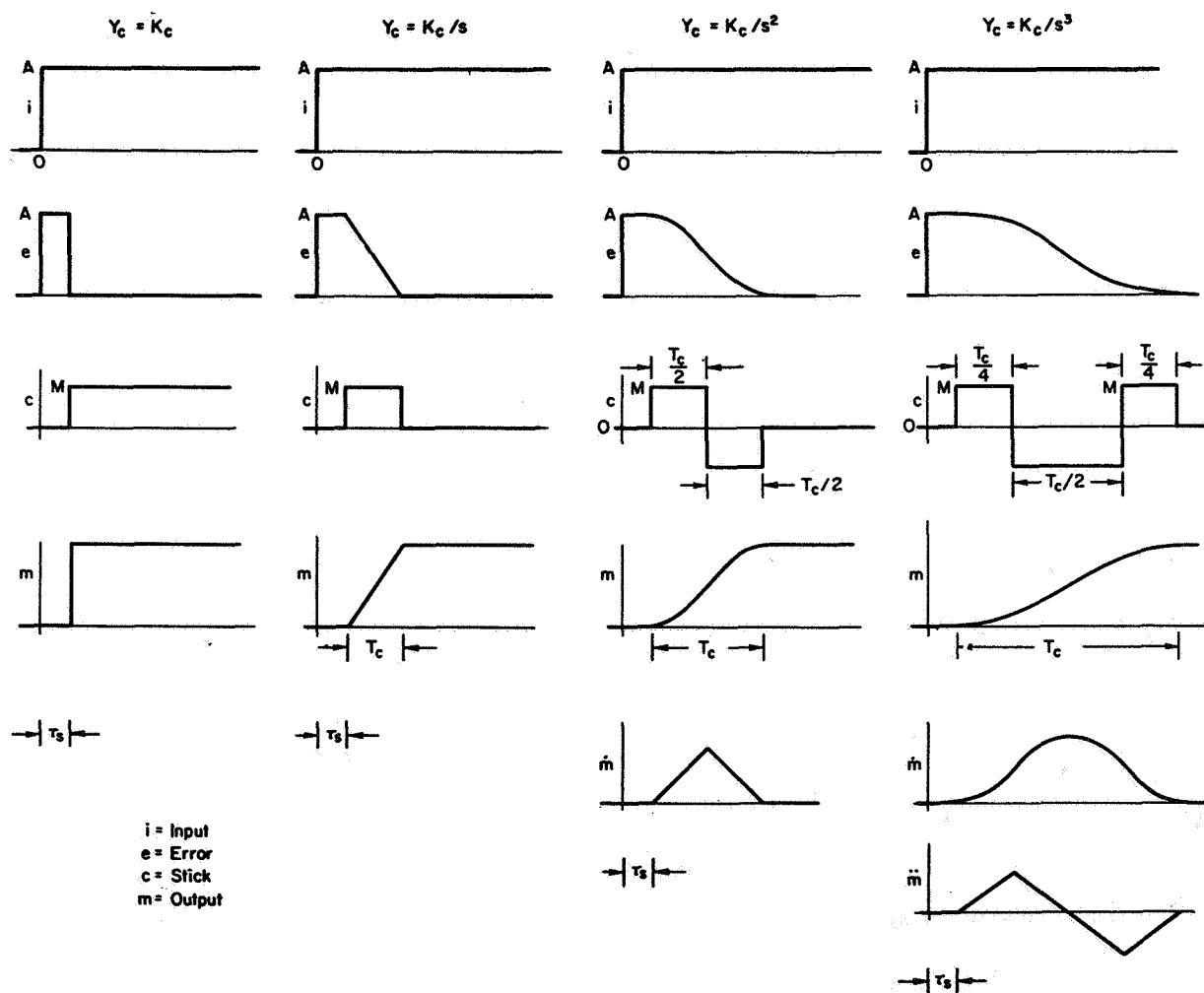


Figure 4.—Ideal time-optimal response characteristics.

TABLE 4. — COMPARISON OF TIME-OPTIMAL INTERPRETATIONS
WITH EXPERIMENTAL RESULTS

CONTROLLED ELEMENT, Y_c	INVARIANCE CONDITION		ERROR* FROM TIME - OPTIMAL
	Optimal	Experimental	
2.86	A	$(M)_o$	$(M)_e$
	8/7	0.4	0.36
	10/7	0.5	0.40
7.1/s	A	$(T_c M)_o$	$(T_c M)_e$
	6/7	0.120	0.130
	8/7	0.161	0.148
	10/7	0.201	0.178
14.3/s ²	A	$(T_c M^{1/2})_o$	$(T_c M^{1/2})_e$
	6/7	0.490	0.488
	8/7	0.564	0.568
	10/7	0.632	0.707
3.6/s ³	A	$(T_c M^{1/3})_o$	$(T_c M^{1/3})_e$
	8/7	2.17	2.04
	10/7	2.33	2.40

$$\text{*Percent error} = \left| \frac{\text{Optimal} - \text{experimental}}{\text{Optimal}} \right| \times 100$$

do affect terminal error rate and acceleration. Perhaps the operator's strategy is to reduce $e, \dot{e}, \dots, e^{(n-1)}$ to small values with the feedforward, rather than to try and make them exactly zero. The quasi-linear mode is used once the error-phase trajectory enters a region near the zero state. Another possibility is that the operator trades off time optimality to minimize some other secondary performance index. It is worth noting in table 2 that the stick amplitudes and pulse durations for different controlled elements seem to be roughly the same. This may correlate with the selection of the particular controlled element gains on a "best" subjective opinion basis.

CONCLUSIONS

A time-optimal control model is one possible idealization of the feedforward step-response path of the dual-mode, mode-switching model for the operator. The complete model is presented in figure 5, as one explanation of available data. The quasi-linear path is the usual operator describing function for compensatory steady-state tracking of random

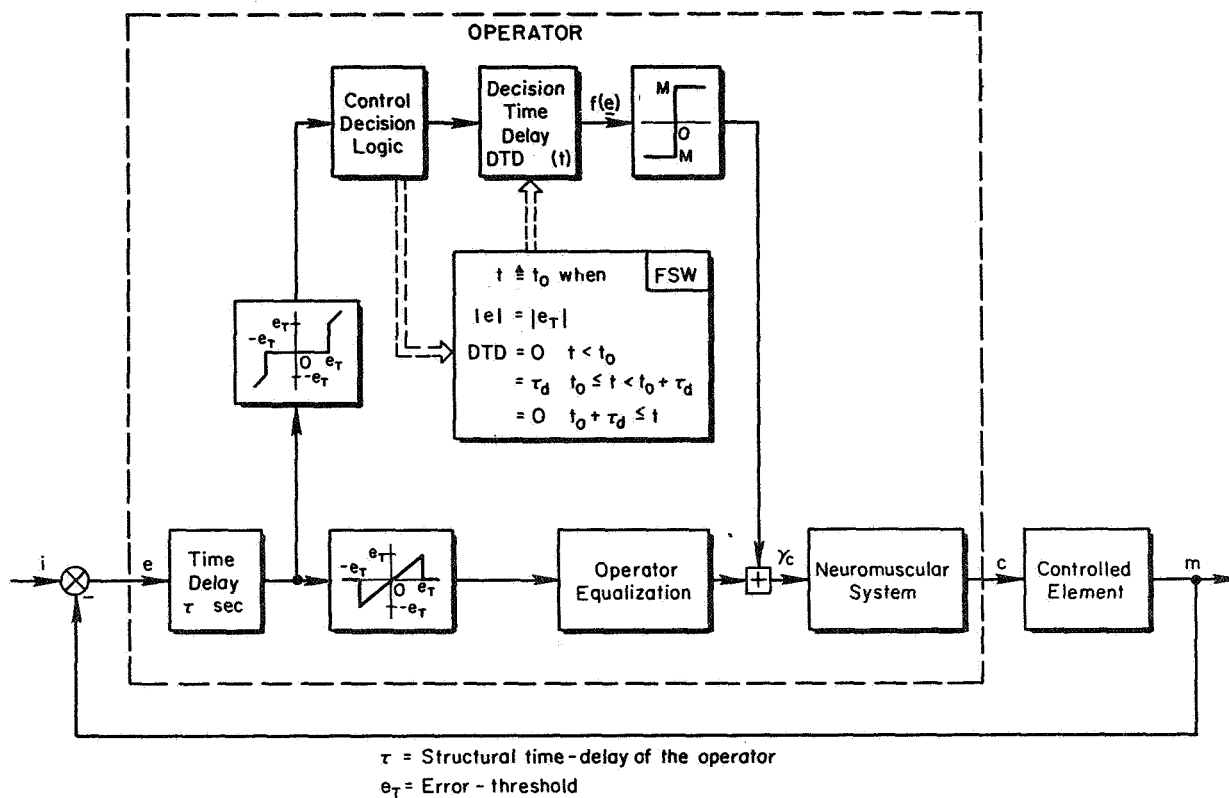


Figure 5.—Dual-mode controller model.

inputs. The feedforward parallel path represents the control-plus-decision model of the operator in response to step inputs. The nonlinear-error-sensing blocks in figure 5 automatically route the error signal through the appropriate channel, based upon its magnitude.

The control logic for each different controlled element and as a function of the error state \underline{e} [$\underline{e} = \text{col}(e, \dot{e}, \dots)$] is given in table 5 for time-optimal response. Note that M , the constraint on the control input, is some function of the step-input height, controlled element gain, and its order.

The decision-logic behaves like a function switch (FSW) and accounts for the initial increase in the time delay (beyond that due to quasi-linear tracking) in response to a step input. Thus, the model of figure 5 serves as one possible explanation of operator behavior in response to random plus-step inputs.

The proposed model is another step toward evolving composite operator models, good for both random and transient inputs. It is an idealization; its optimality and sensitivity need further examination, and other performance indices may be appropriate as well. Nevertheless, the time-optimal form does provide a useful and reasonably accurate description.

TABLE 5.—CONTROL LOGIC FOR VARIOUS CONTROLLED ELEMENTS

CONTROLLED ELEMENT, Y_c	CONTROL LOGIC, $f(e)$
K_c	$(A/MK_c)e(t)$
K_c/s	$e(t)$
K_c/s^2	$[\dot{e} + \sqrt{2MK_c} e \operatorname{sgn} e]$
K_c/s^3	$e + (1/3)\ddot{e}^3 + W\dot{e}\ddot{e} + W[(1/2)\ddot{e}^2 + W\dot{e}]^{3/2}$ $W = +1 \text{ for } [\dot{e} + (1/2)\ddot{e} \ddot{e}] > 0$ $= -1 \text{ for } [\dot{e} + (1/2)\ddot{e} \ddot{e}] < 0$

REFERENCES

1. McRuer, D. T.; Hofmann, L. G.; Jex, H. R.; et al.: New Approaches to Human-Pilot/Vehicle Dynamic Analysis. AFFDL-TR-67-150, Oct. 1967.
2. Elkind, J. I.; Kelly, J. A.; and Payne, R. A.: Adaptive Characteristics of the Human Controller in Systems Having Complex Dynamics. IEEE Proc. Fifth Natl. Symp. HFE, May 1964, pp. 143-159.
3. Young, L. R.: A Sampled-Data Model for Eye Movements. Sc.D. Thesis, MIT, Dept. Aeron. Engr., June 1962.
4. Lemay, L. P.; and Westcott, J. H.: The Simulation of Human Operator Tracking Using an Intermittent Model. Paper, IRE Intern. Cong. HFE (Long Beach, Calif.), May 3-4, 1962.
5. Tsibulevskii, I. E.: Operator Delay in Processing Visual Signals. Avtomatika i Telemekhanika, vol. 23, no. 11, Nov. 1962, pp. 1513-1526.
6. Graham, D.; and McRuer, D.: Analysis of Nonlinear Control Systems, John Wiley & Sons, Inc., 1961.

22. State-Space Models of Remote Manipulation Tasks

*Daniel E. Whitney and Thomas B. Sheridan
Massachusetts Institute of Technology*

The state-space method has been shown to be capable of describing a wide variety of the logical and physical constraints which comprise manipulation. Using it, we can plan some non-trivial tasks and gain insight into the nature of tasks and commands. Combined with a flexible input language and carefully composed execution routines, it can provide a basis for supervisory controlled manipulation.

INTRODUCTION

Remote manipulation is usually difficult even if the human operator is close to his work because, typically, there is meager feedback and the apparatus is clumsy and hard to control. Add to this a significant time delay, and efficient manipulation becomes almost impossible. It has been shown (ref. 3) that in the absence of force feedback, the operator may overcome time delay by adopting a "move-and-wait" strategy. In the absence of time delay, force feedback has been shown to improve the performance of manually controlled manipulation (ref. 6). However, force feedback actually hinders manipulation in the presence of time delay unless the operator deliberately ignores the force feedback precisely at those moments when it is most useful, prior to and during unexpected collisions. To accomplish this, the operator must again adopt the move-and-wait strategy, periodically removing his hand from the controls (ref. 4).

This paper presents work on a different approach to this problem. Called supervisory controlled manipulation, the idea is to increase the manipulator's own intelligence to the point where direct manual control by the human operator is unnecessary throughout the major parts of most tasks. Instead, the operator gives commands of a verbal nature to a computer which plans manipulator motions and sends the plan to a smaller computer located near the manipulator. The small computer executes the plan, providing simple feedback analysis and reflex actions (under no time delay) and asking for help when the execution hits a snag. Delay remains an inconvenience, but in principle, the amount of delay has no effect on the actual time taken to form plans and put them into action. Figure 1 depicts the supervisory controlled manipulator concept.

The main advantage of such a system is that it gives the operator the kind of decision and pattern recognition tasks for which he is well suited, leaving to the computers the routine work of detailed motion planning and execution which is so difficult for the distant operator.

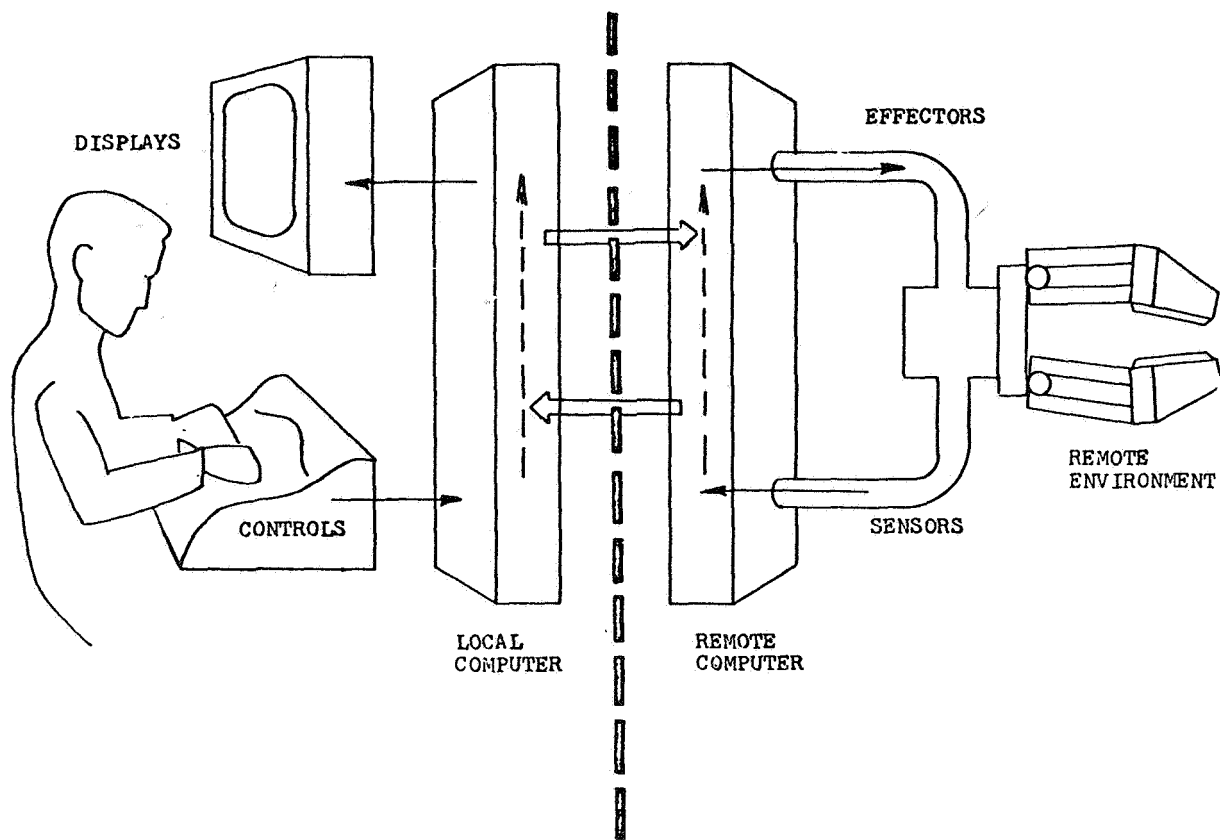


Figure 1.—Schematic diagram of supervisory controlled remote manipulator.

The creation of a supervisory controlled manipulator poses three main challenges:

- (1) How is the system equipped in order to understand what the operator wants done?
- (2) How is the system to translate the operator's desire into a plan of action relevant to the task environment and capable of achieving the stated goal?
- (3) How is the plan executed?

This paper concentrates on the first two of these challenges. The third has been studied (but by no means exhausted) by Ernst (ref. 2).

TASKS AND COMMANDS

Designers of manually controlled manipulation systems attempt to give the operator the feeling that he is at the remote site. The operator of such a system must attend to every detail of task planning and execution. The handicaps of limited touch and vision feedback and time delay cause even the smallest detail to become a project in its own right. Pictorially, we may show such a project at the focus of a hierarchy of decision functions (fig. 2).

The computers in a supervisory controlled manipulator stand between the operator and the manipulator, both in a physical sense and also in an "intellectual" sense, so that they occupy a place in the decision diagram. Depending on how capable we can make this system, the computer might occupy one of several levels, or may occupy different levels at different times during normal operation, as seen in figure 3.

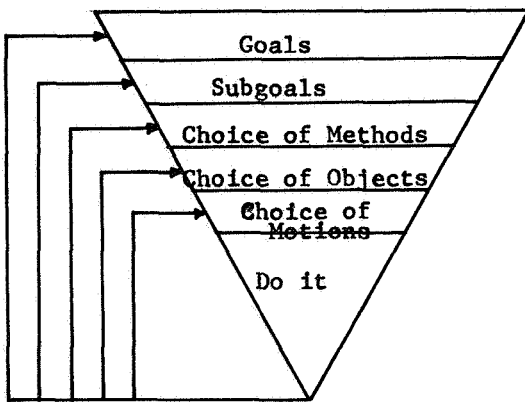


Figure 2.—Pictorial representation of decision levels in manual control of remote manipulation.

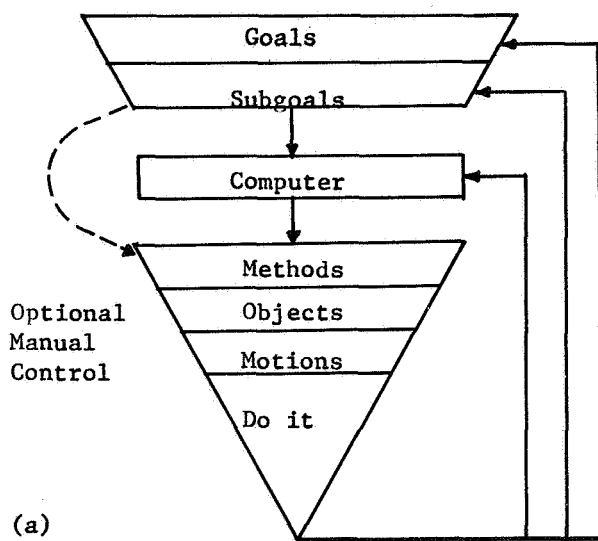
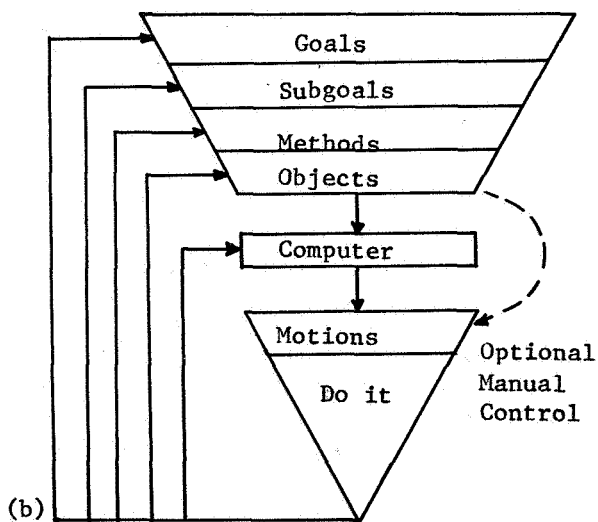


Figure 3.—Two versions of decision levels in supervisory control of remote manipulation (a) more capable (b) less capable.



A major difference between the schemes in figure 2 on one hand and figure 3 on the other is that in the latter the operator passes the work on to the computer at a level far enough above the little details so that he can think broadly about the task: in figure 3(a), for example, he can think about several subgoals, being relieved of the need to concentrate continually on manual details. This means that he can either supervise several manipulators at once or can think about future tasks for one manipulator. In either figure 3(a) or figure 3(b), he may give goal-oriented commands to the computer which generates subgoal-oriented commands which are directly intelligible to the manipulator's prime movers. By contrast, the operator of a manually controlled manipulator may be said to be giving method commands at a very detailed level.

This means that the supervisory controlled manipulator cannot be a mere servo, but must be able to plan ahead, at least a little; analyze feedback; organize and store information about the task site; and be able to receive and understand commands like, "Put the pencil on the table, next to the eraser." (We assume that commands are coded in some way to permit their input. We do not imply that analysis of English syntax is either desirable or necessary for implementation.) In such commands, the objects, places, and actions are referred to by name.

We define a manipulation task as an alteration in the positions and orientations of objects and effectors (jaws, tools, etc.) or, more generally, as an alteration in the geometric configuration of the task site. Thus we specifically exclude such activities as bouncing a ball or balancing a stick on end. By ignoring velocities and accelerations, we may concentrate on the geometric constraints fundamental to manipulation: obstacle avoidance, rendezvous of jaws and objects to be grasped or pushed, rendezvous of a jaw-borne object and its destination for pushing, releasing, pouring, inserting, and so on. (Granted a jaw-borne object has a velocity: nevertheless what is important about carrying is that jaws and object maintain a fixed geometric relation to each other throughout.)

If tasks are geometric, then it makes sense to break them up into static submotions which are general enough so that only a few are needed to describe many tasks. Then, with these few submotions, taken in the right quantity and order, a wide variety of tasks can be planned and, hopefully, accomplished. To be precise, call the submotions atomic commands. For example:

Move jaws left 1 inch
Move jaws right 1 inch
Open jaws
Close jaws

In what follows we shall describe a method by which a task site may be modeled. The operator may request a task, and the computer can devise a sequence of atomic commands which, to the degree of precision of the task model, can accomplish the task.

THE STATE-SPACE MODEL

The problem we have posed is to find a sequence of finite elements which has a particular property. We may dispense with enumeration of the possibilities, since there are far too many, and far too few of them are worth considering.

Instead, let us recognize that the result of accomplishing one atomic command is a new configuration of the task site, a configuration which differs from its predecessor in exactly one well-defined way: the jaws have been opened, or the jaws are now displaced 1 inch left from their previous location, and so on. Let a point in an abstract space represent one such configuration. Call this point a state of the task. Next to it put another state. The two states will be called neighbors if executing one atomic command carries the task from one

state to the other. We show this by connecting the two neighboring states by a line bearing an arrow to show in which direction the transition may be made. (In most cases, an atomic command will be invertible so that transitions in both directions are allowed; the arrow is then omitted.) The totality of such points, some connected by lines, is called the state space. Its points represent all the configurations which the task site can assume as a result of the execution of arbitrary strings of atomic commands. We may then say that paths through the space represent strings of atomic commands which make coherent (though not necessarily purposeful or efficient) changes in the task site. For example, consider the task site in figure 4. The jaws may move from point to point along the line, open and close. Given the atomic command set above, the state space corresponding to figure 4 is shown in figure 5.

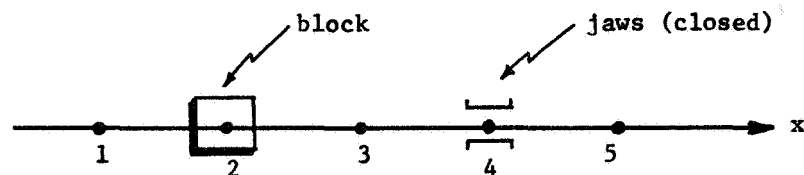


Figure 4. —One-dimensional task site.

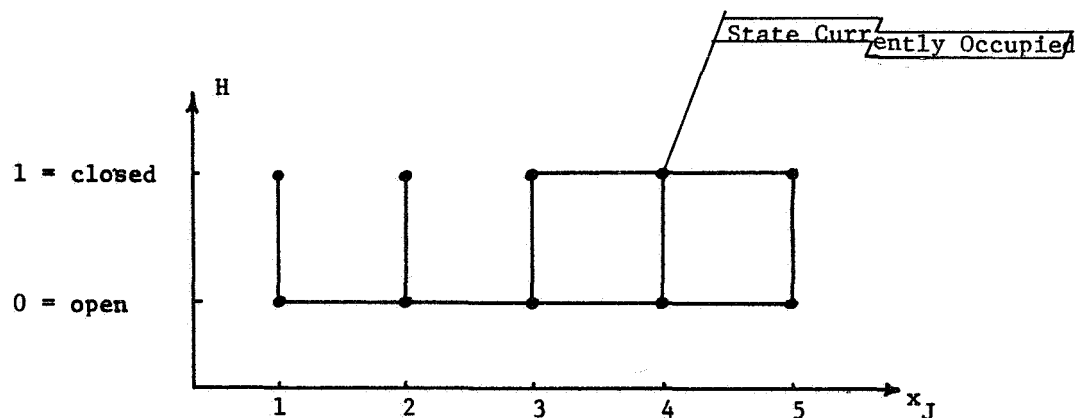


Figure 5. —State space corresponding to figure 4.

The coordinates on the axes are x_J , jaw position, and H , jaw status. x_J and H are called the state variables. The configuration of figure 4 (jaws in location 4, closed) is represented by the flag at state $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ in figure 5. The vertical lines indicate that the jaws may open or close at any x , while the horizontal lines indicate allowed jaw movement. Movement into $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ from $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is forbidden since this would involve a collision between jaws and object. Thus $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ are not neighbors in the sense defined above.

If we wish the jaws to move to location $x=1$, jaws closed, we so indicate by asking the computer to take the task site to state $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. (Naturally we want the computer to derive by itself the fact that the jaws must straddle the object on the way.) We therefore want the

computer to find a satisfactory path in state space from state $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ to state $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Such a path, if it exists, can be translated immediately into a string of atomic commands suitable for accomplishing the task, since the path tells the sequence of neighboring intermediate configurations through which the task site should pass on the way to the desired configuration. There are countless possible paths, most of which go nowhere purposive. But a shortest path cannot go nowhere and, in particular, cannot loop. So let us find shortest paths.

Many algorithms are available for finding shortest paths in networks, among them dynamic programming (ref. 1), Ford's algorithm (ref. 5), and the Hart-Nilsson-Raphael algorithm (ref. 7). Of more interest to us is the general interpretation we can give to "shortest": A path may be short in time, fuel, risk, lack of information, or some (normalized) combination of these, for example. By specifying the dimension, magnitude, and distribution of the "lengths" of lines between points in the state space, the operator can impose his decisions quite effectively. The arrangement of lengths shown in figure 6 results in the path indicated by the arrows.

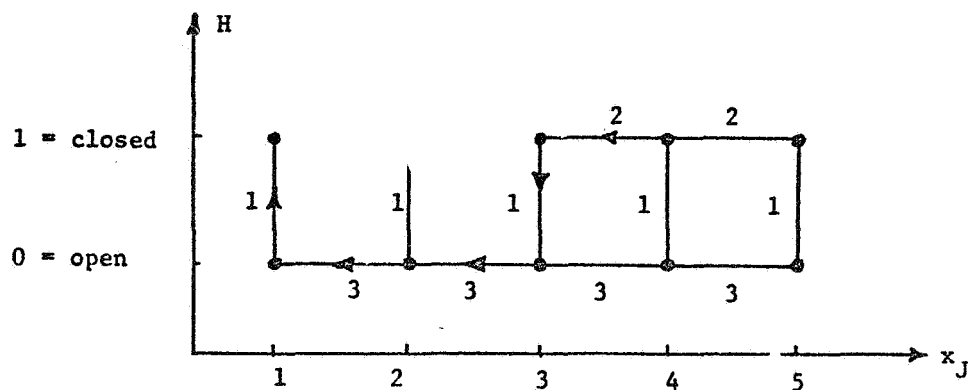


Figure 6.—State space with assigned lengths and the shortest path from $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

The corresponding work plan is—

Move jaws left 1 inch
Open jaws
Move jaws left 1 inch
Move jaws left 1 inch
Close jaws

A path which allows the jaws to grasp the object is shown in figure 7. The corresponding path for the case where the object is in location $x=1$ is shown in a new state space in figure 8. By making the object's location a new state variable y_0 we may represent carrying and pushing in a larger state space (fig. 9). Note that pushing is not expressed as a transition analogous to that $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ in figures 5, 6, and 7, for example.

This type of state transition is still forbidden since it tells nothing about what happens to the pushed object. Pushing is properly expressed in figure 9 as a variation of carrying.

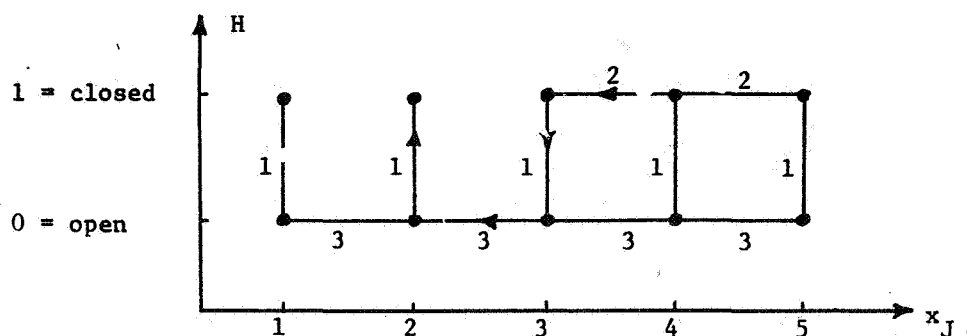


Figure 7. —Path which represents grasping the object.

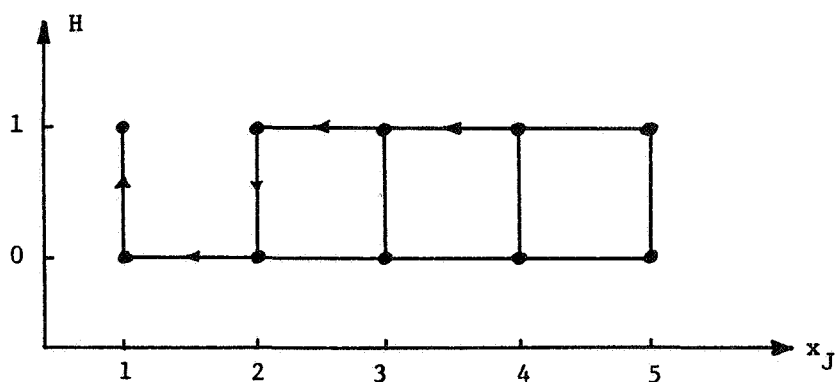


Figure 8. —Path for grasping the object in location 1.

A MORE COMPLEX EXAMPLE

Suppose we wish to move a long thin spar through a crowded two-dimensional environment. Here the interactions between the environment and the spar's position and orientation are of the most interest. To map motions of the spar once it is grasped by a rotating pair of jaws, we choose state variables

x = x coordinate of spar, $1 \leq x \leq 5$

y = y coordinate of spar, $1 \leq y \leq 3$

α = orientation of spar

$\alpha = \begin{cases} 0, & \text{if spar is parallel to } x\text{-axis} \\ 1, & \text{if spar is parallel to } y\text{-axis} \end{cases}$

The allowed commands are—

Move $\pm x$ direction one unit

Move $\pm y$ direction one unit

Rotate 90°

Thus both position and orientation of the spar are quantized.

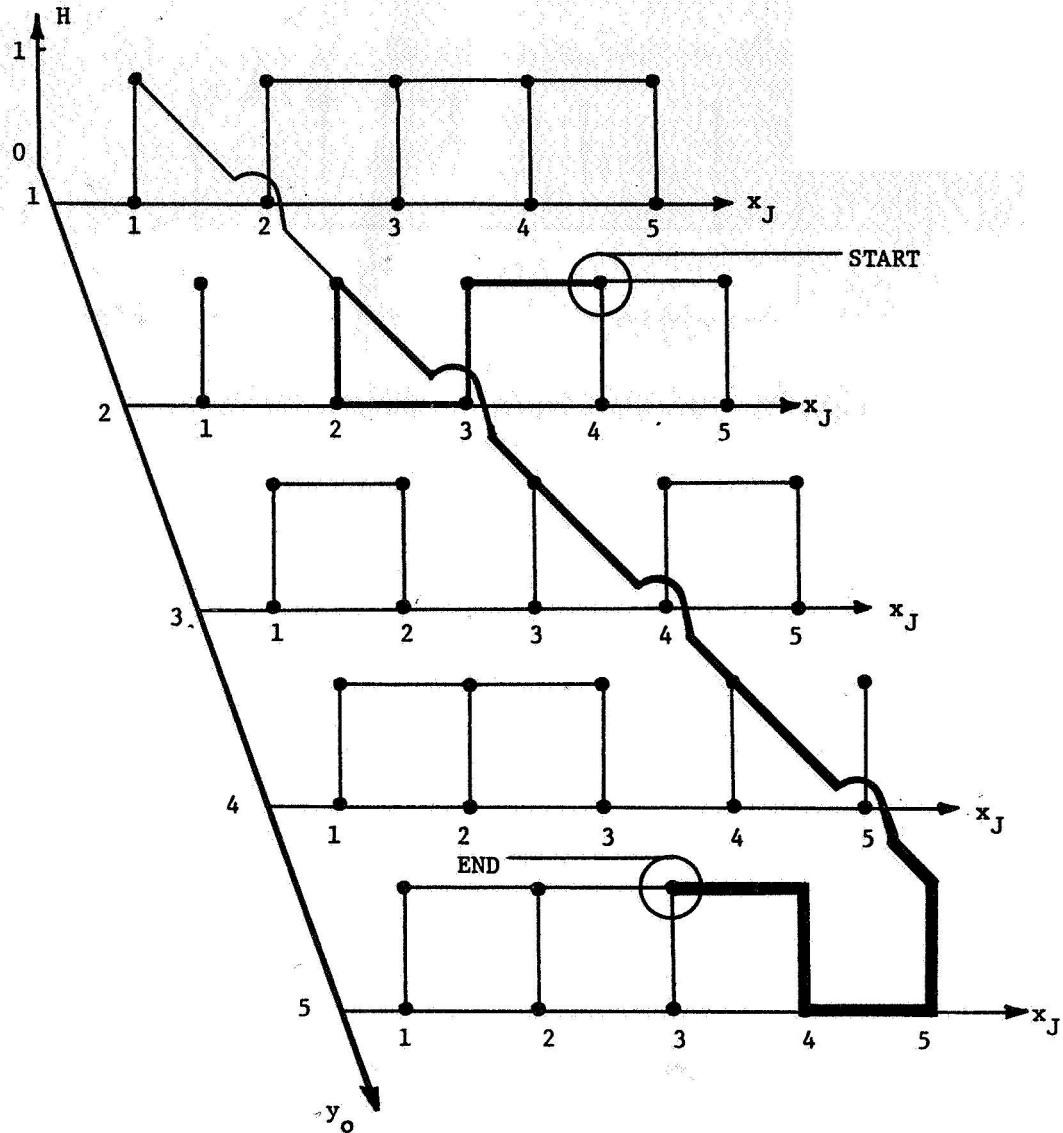


Figure 9.—State space suitable for all carrying tasks.

The physical space is shown in figure 10. Walls are shown as open rectangles, while the two possible orientations of the spar are shown by cross lines at each possible position. The challenge is provided by the doorways, which allow the spar to pass axially but not athwart. This constraint and the presence of walls are shown in the state space by deleting the lines corresponding to the forbidden transitions.

The state space appears in figure 11. We assume for illustration that each "move" is of length 2, each "rotate" of length 3. Let the spar be initially at location (2, 2), oriented parallel to the y -axis, and say we want it moved to (3, 3), ending up oriented parallel to the x -axis. Then the initial state is $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ and the final state is $\begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix}$. These are marked "Start" and "End," respectively, on figure 11.

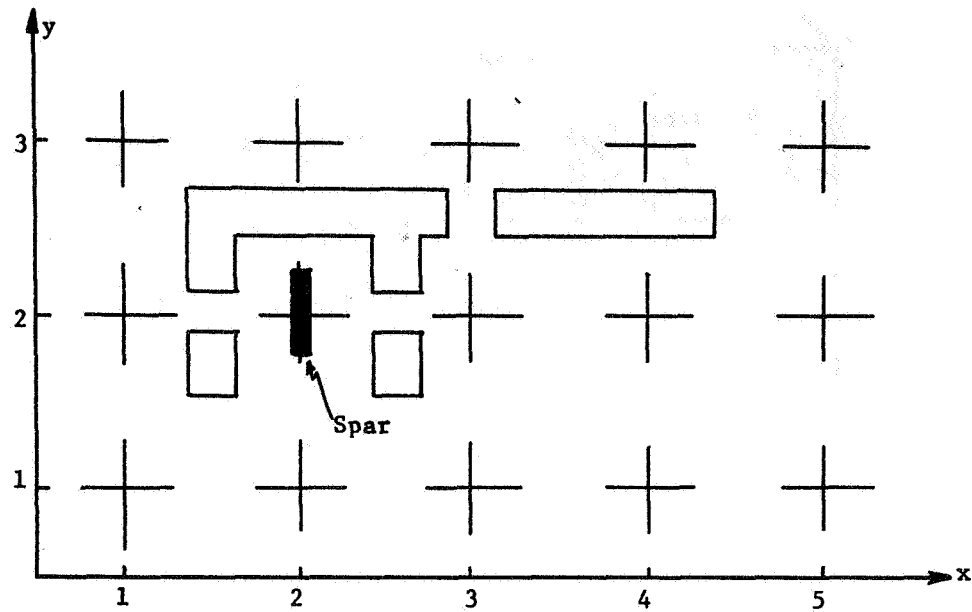


Figure 10. —Physical space for the spar problem.

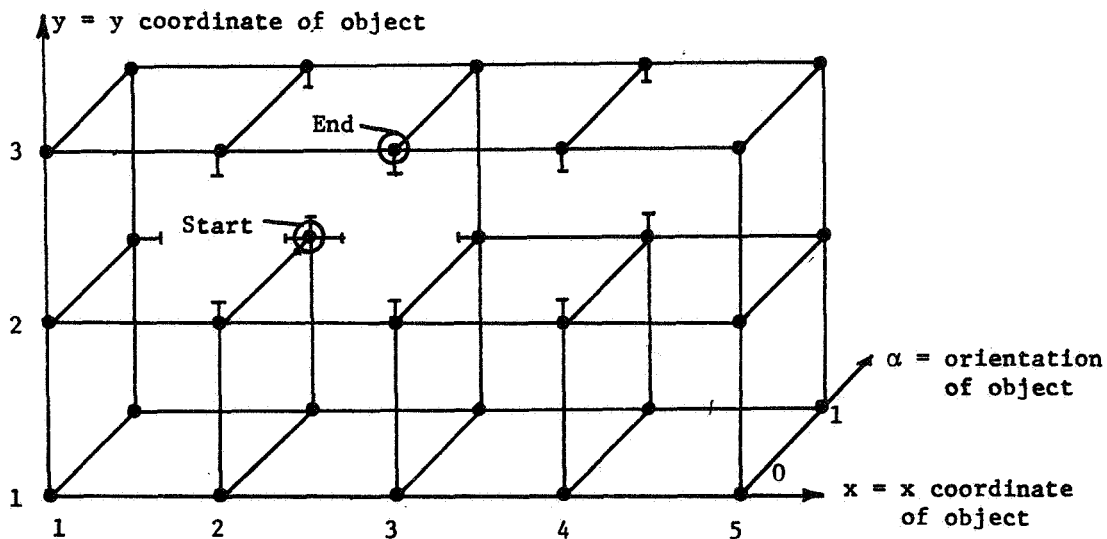


Figure 11. —Graph of the problem in figure 10.

There are two equal-length solution paths, shown in figure 12 and visualized on a sketch of the task site in figure 13. These paths do not "look like" the most direct route. Closer examination, however, reveals that these paths, by initially moving the object away from the final state, are able to save two rotations by spending a little more distance. Again, if we read a solution path, we get a list of the required moves and rotates in the correct order. A more general solution to this problem, which includes grasping and releasing the spar in arbitrary (quantized) positions and orientations, may be found in chapter V of reference 8.

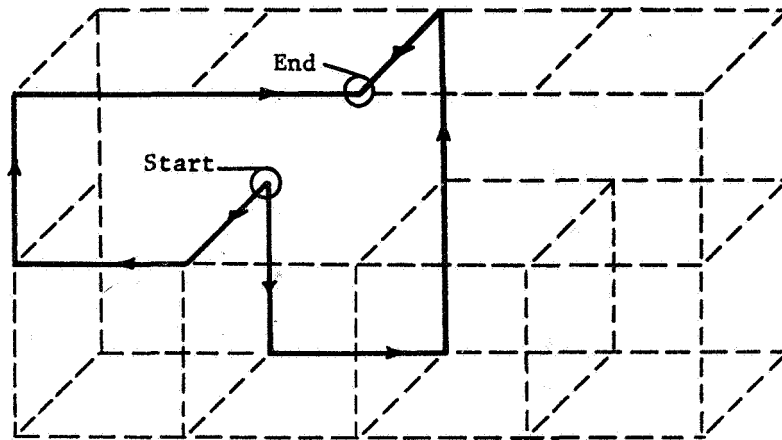


Figure 12. —Solution to the graph of figure 11.

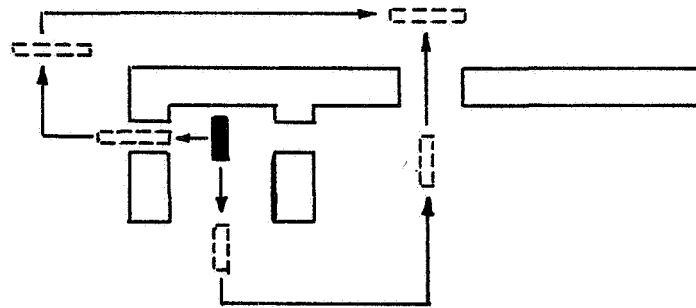


Figure 13. — Visualization of solution paths from figure 12.

EXTENDING THE POWER OF THE STATE-SPACE METHOD

Now we have reduced some basic manipulation tasks to shortest-path problems. In this section we show how to make use of the basic pick-up-and-carry capability demonstrated in figure 9 to program more complex tasks, preserving the operator's ability to shape the general features of the result without having to specify details.

We think of commands as functions whose arguments (specified by the operator) may be place or object names, and whose values are paths in one or more related state spaces. (The details of evaluation of these functions are described more fully in ref. 8.) The basic function "Take" is already established by the existence of state spaces such as figure 9. Formally:

$$\text{Take (A, X, Y)} = \text{Carry object A to location X, leaving the jaws in location Y.} \quad (1)$$

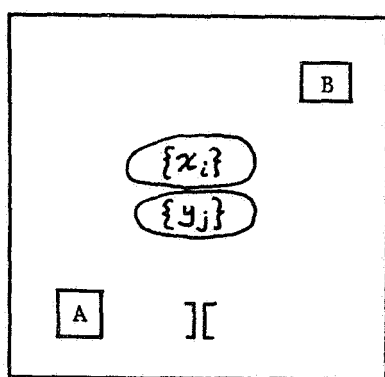
To exchange objects A and B, we may use Take to define Switch:

$$\begin{aligned} \text{Switch (A, B, X, Y)} = & \text{Take (A, X, Y)} \\ & + \text{Take (B, A', Y)} \\ & + \text{Take (A, B', Y)} \end{aligned} \quad (2)$$

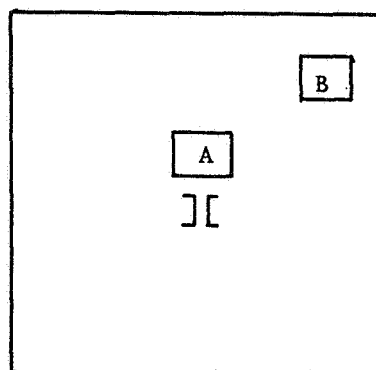
in which "+" means "followed by," X is a location between A and B (chosen by the operator), Y is a location near X, A' is A's old location and B' is B's old location. Switch may be "optimized" if we allow the operator to define fuzzy areas \bar{X} and \bar{Y} in which points X and Y, respectively, are to be sought by the computer so as to achieve minimum total task length:

$$\text{Switch } (A, B, \bar{X}, \bar{Y}) = \min \left\{ \begin{array}{l} \text{Take } (A, x_i, y_j) \\ \{x_i\} + \text{Take } (B, A', y_j) \\ \{y_j\} + \text{Take } (A, B', y_j) \end{array} \right\} \quad (3)$$

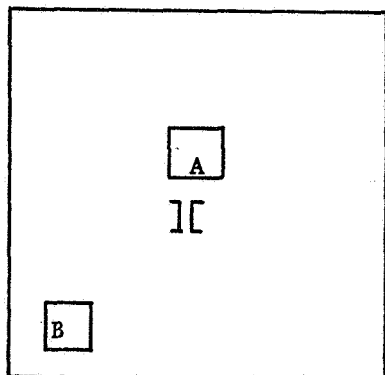
in which $\{x_i\}$ and $\{y_j\}$ denote the sets of locations in regions \bar{X} and \bar{Y} , respectively. Four stages in the planning of this task are shown in figure 14. The advantage of (3) over (2) is



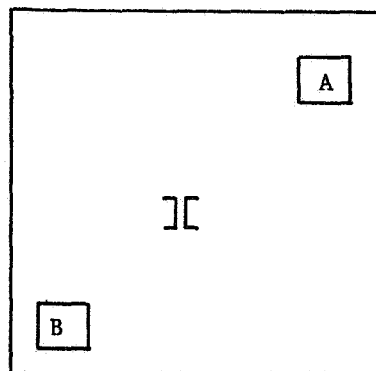
I. Starting Configuration.



II. Configuration after taking A to the best x_i , leaving jaws in the best y_j for the whole task.



III. After taking B to A's old location, leaving jaws in same y_j as in step II.



IV. After taking A to B's old location, leaving jaws at same y_j as in steps II and III.

Figure 14. — Four stages of plan for switch (A, B) with optimal stopping places for A and jaws deduced automatically.

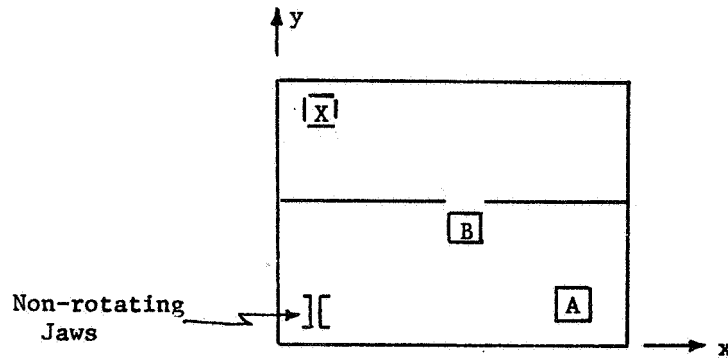


Figure 15.—Physical space for the blocked doorway problem.

that the operator need not have precise knowledge concerning \bar{X} and \bar{Y} in order to request a switch. \bar{X} and \bar{Y} correspond to the statement "over there somewhere," a convenient sort of statement to be able to make.

Function definition is an obvious way of increasing the manipulative capabilities of a computer-manipulator. We may build layer on layer, using the existing functions to define new ones. We have a way of telling the computer "how" to do new tasks, such as assembling involved structures or making complex selection and ordering of subgoals. As an example of the latter, consider the task site in figure 15.

Several objects block a narrow doorway. We wish to move object A to location X. It is possible to define a function which enables the computer to decide, on the basis of minimum total task length, which blockers to move, the order in which to move them, and where to move them, in order to get A to X. The operator, when defining the function, says in effect, "These are the blockers, here is a region where they might be put; move A to X." This problem is discussed more fully in chapter V of reference 8.

REFERENCES

1. Bellman, R. E.; and Dreyfus, S. E.: Applied Dynamic Programming. Princeton Univ. Press, 1962.
2. Ernst, H. A.: A Computer-Operated Mechanical Hand. Sc.D. Thesis, MIT, Dept. Elect. Engin., Dec. 1961.
3. Ferrell, W. R.: Remote Manipulation with Transmission Delay. IEEE Transactions on Human Factors in Electronics, HFE-6 (1), Sept. 1965, pp 24-32.
4. Ferrell, W. R.: Delayed Force Feedback. Human Factors, Oct. 1966, pp 449-455.
5. Ford, L. R.: Network Flow Theory. Rand Corp. Paper P-923, Aug. 14, 1956.
6. Goertz, R. C.: Manipulators Used for Handling Radioactive Materials. Human Factors in Technology, Ch. 27, McGraw-Hill, 1963.
7. Hart, P. E.; Nilsson, N. J.; and Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. Stanford Research Institute, unpublished memo, June 1967.
8. Whitney, D. E.: State Space Models of Remote Manipulation Tasks. MIT, Dept. Mech. Engin., Ph.D. Thesis, Jan. 1968.

V. SYSTEM IDENTIFICATION, SIGNAL ANALYSIS, AND PERFORMANCE MEASUREMENT

23. Identification of Sampling Intervals in Sampled Data Models of Human Operator*

*G. A. Bekey and C. B. Neal
University of Southern California*

This paper develops two methods for the determination of an unknown sampling interval in closed-loop sampled-data systems. Both methods are based on a priori knowledge of the structure of the system to be identified and require that measurements be performed only on the system input and its (continuous) output.

INTRODUCTION

This paper is concerned with the development of systematic procedures for the identification of unknown sampling frequencies in closed-loop linear discrete-time systems. While a number of techniques for parameter identification in both continuous and discrete systems are well known (e.g., gradient methods, random search, relaxation, quasi-linearization, etc.), these techniques have not been applied to systems where the sampling frequency is one of the parameters to be identified. The problem is of particular interest in closed-loop systems with error sampling, where only the continuous input and the continuous output are observable.

The techniques developed here are of particular interest in the study of man-machine systems and certain biological systems. The human operator in manual-control systems has been hypothesized to behave as a discrete-time system. However, both his visual input and hand-controller output are continuous.

STATEMENT OF THE PROBLEM

Consider the basic identification problem illustrated by fig. 1, showing a system with an unknown sampling frequency f_x (or sampling interval $T_x = 1/f_x$) and an adjustable sampling interval T_a . The following assumptions are made:

- (1) The structure of the system is known a priori; therefore, the data holds, and dynamics of the system and model agree exactly. Thus, the only unknown factor is the sampling interval T_x . Later, the requirement that system and model agree will be relaxed.
- (2) The system and model are noise-free.

*This research was supported in part by the National Aeronautics and Space Administration under NASA Grant NGR 05-018-022.

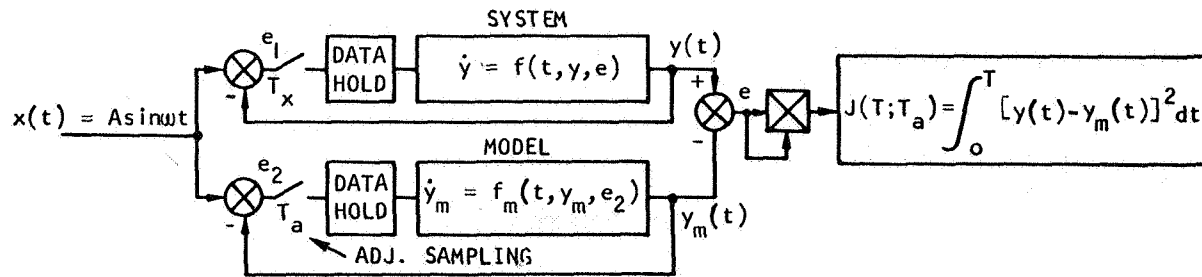


Figure 1. —Identification scheme.

(3) The differential equations of the system and model have unique solutions. This means that if

$$\dot{\bar{y}} = f(t, \bar{y}(t), \bar{u}(t)) \quad (1)$$

is the form of the vector differential equation for both the system and model then both $f(\cdot)$ and $\frac{\partial f_i}{\partial y_i}(\cdot)$ must be continuous and bounded. (In eq. (1), \bar{y} and \bar{f} are r -dimensional vectors, and \bar{u} is an $m \leq r$ vector.)

The identification of T_x will be based on the minimization over T_a of a criterion function, chosen as

$$J(T; T_a) = \int_0^T [y(t; T_x) - y_m(t, T_a)]^2 dt, \quad (y, y_m \text{ scalars}) \quad (2)$$

ANALYTICAL FOUNDATION

For the problem stated in the previous section, the following theorems can be shown to hold:

Theorem I: A necessary and sufficient condition for the identification of a sampling interval T_x is that

$$\min_{T_a} J(T; T_a) = 0 \quad (3)$$

and for this case $T_x = T_a$.

Theorem II: The criterion function J cannot be zero on a T_a interval; i.e., J is zero for one value of T_a only.

Theorem III: If the system and model are linear, then J is strictly convex over T_a . (The implication of this theorem is that convergence to a unique value of T_x may be expected when a convergent computer algorithm is used for the identification.)

IDENTIFICATION OF T_x BY AN ITERATIVE DISCRETE GRADIENT METHOD

This method is based on the computation of the function u_T which provides a measure of the sensitivity of system response to variations in sampling interval T_a (ref. 1).

We assume that the system being identified can be described by the equation

$$\dot{\bar{y}} = \bar{f}(t, \bar{y}(t), e(t), \bar{a}) \quad (4)$$

where \bar{a} is a vector of system parameters,

$$\bar{a} = (a_1, a_2, \dots, a_n, T_x) \quad (5)$$

A model constructed in accordance with the assumptions under "Statement of the Problem" is described by

$$\dot{\bar{y}}_m = \bar{f}(t, \bar{y}_m(t), e(t), \bar{p}) \quad (6)$$

with the parameter vector

$$\bar{p} = (p_1, p_2, \dots, p_n, T_a) \quad (7)$$

The criterion function of equation (2) is evaluated over some $T \gg T_a$. Also, the gradient vector

$$\nabla_{\bar{p}} J(T; \bar{p}) = -2 \int_0^T [y(t; T_x) - y_m(t, T_a)] \nabla_{\bar{p}} y_m(t; T_a) dt \quad (8)$$

is computed. Then the iterative gradient adjustment follows according to

$$\bar{p}_{i+1} = \bar{p}_i - K_i \nabla_{\bar{p}_i} J(T; \bar{p}_i) \quad (9)$$

Now, for sampled-data systems, the vector function $\nabla_{\bar{p}} y_m(t)$ is most conveniently expressed by using a difference equation representation of the solution of the differential equation (ref. 2). Corresponding to the differential equation of the model of equation (6) there is a difference equation for the solution at sampling instants

$$\bar{y}_m[(n+1)T_a] = \bar{f}_m[nT_a, \bar{y}_m(nT_a), e(nT_a), \bar{p}] \quad (10)$$

Define the sampling interval global sensitivity function (ref. 1) by

$$\bar{u}_{T_a}(nT_a) = \left. \frac{\partial \bar{y}_m(t)}{\partial T_a} \right|_{t=nT_a} = \lim_{\Delta T_a \rightarrow 0} \left[\frac{\bar{y}_m[n(T_a + \Delta T_a)] - \bar{y}_m(nT_a)}{\Delta T_a} \right] \quad (11)$$

where $n \Delta T_a \ll T_a$. This function is obtained by solution of appropriate sensitivity difference equations (ref. 1).

Example.—The identification scheme of figure 1 is used with system K_1/s ; model K_3/s . Zero-order data holds are used in both loops. The output of the model loop at standing instants is

$$y_m[(n+1)T_a] = y_m(nT_a) + T_a K_3 [x(nT_a) - y(nT_a)] \quad (12)$$

The sampling interval global sensitivity difference equation, using equation (11) and equation (12) is

$$u_T[(n+1)T_a] = [1 - T_a K_3] u_T(nT_a) + T_a K_3 \left[\frac{1}{T_a} (t \dot{x}(t)) \right]_{t=nT_a} + T_a K_3 \left[\frac{x(nT_a) - y(nT_a)}{T_a} \right] \quad (13)$$

The sensitivity difference equation for the gain parameter is

$$u_K[(n+1)T_a] = (1 - T_a K_3) u_K(nT_a) + T K_3 [(x(nT_a) - y(nT_a))/K_3] \quad (14)$$

Equations (13) and (14) are then substituted in equation (8), and the corrected parameter vector is then obtained from equation (9). The results of the gradient search for T_x alone, and for both T_x and K_1 are shown in figure 2.

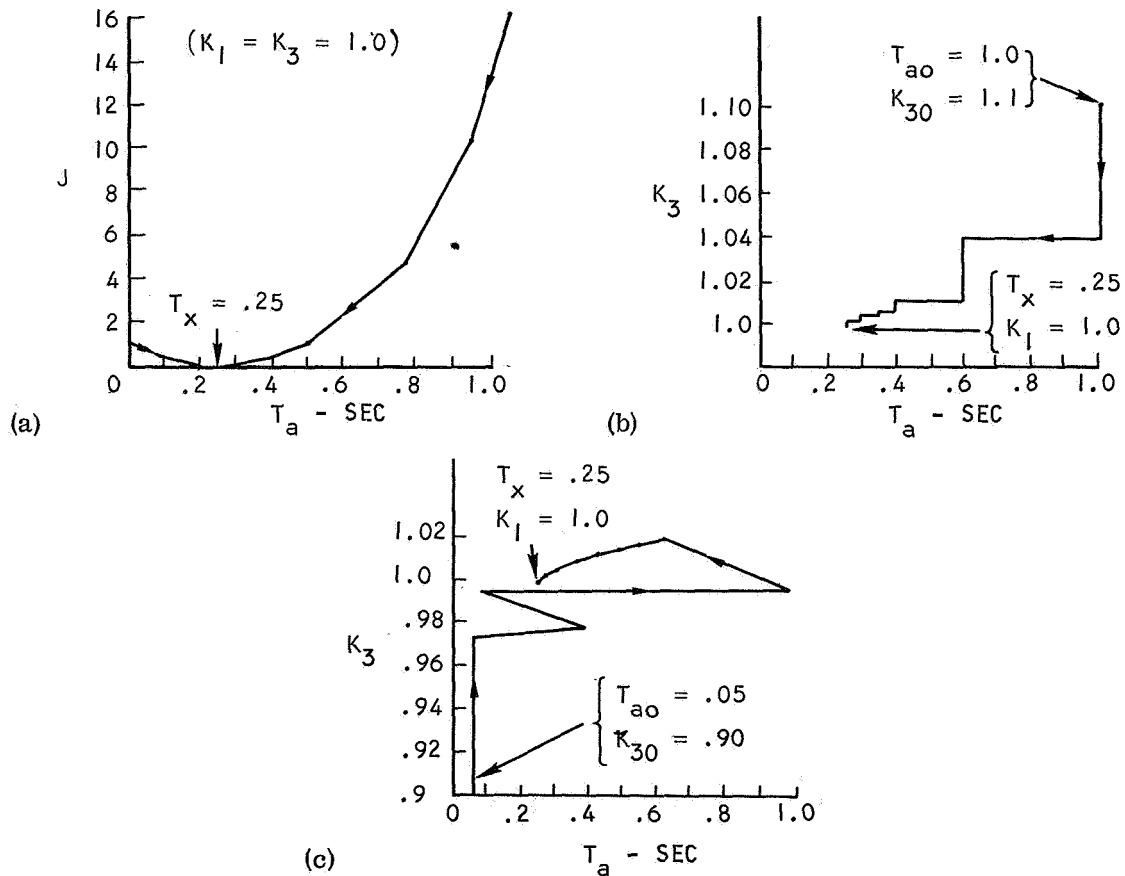


Figure 2. —Results of gradient search for T_x and T_x and K_1 . (a) Iterative search for T_x alone. (b) Alternating search for T_x and K_1 . (c) Simultaneous search for T_x and K_1 .

PROGRAMED SEARCH FOR T_x

An alternative to gradient methods is the use of programed search for obtaining the best estimate of T_x . Again, the modeling scheme of figure 1 applies. Linear systems were chosen because their solutions from given initial conditions are unique. Programed search was applied successfully to the system-model combinations of table 1 (see figs. 3, 4, and 5). Data holds were both zero-order. The experimental results indicate that when the model form is correctly chosen the presence of sampling can be detected quite readily by the minimum in the criterion-function (J) curve. In the case where the model form is imperfectly known (in violation of assumption 1 above), experimental results indicate the J-curve may have no discernable point of inflection, as was the case with the model of experiment 3 in table 1. The lack of a minimum can be taken as an indication that other models should be tried.

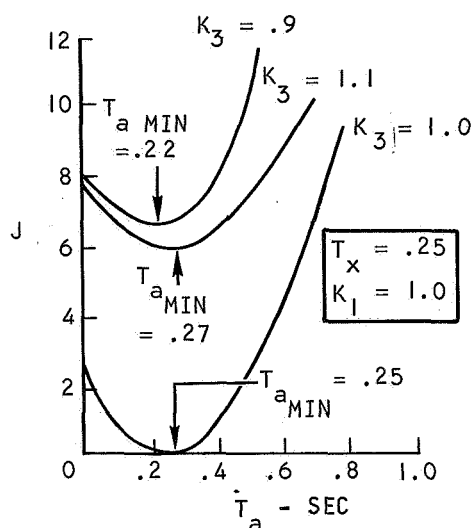


Figure 3.

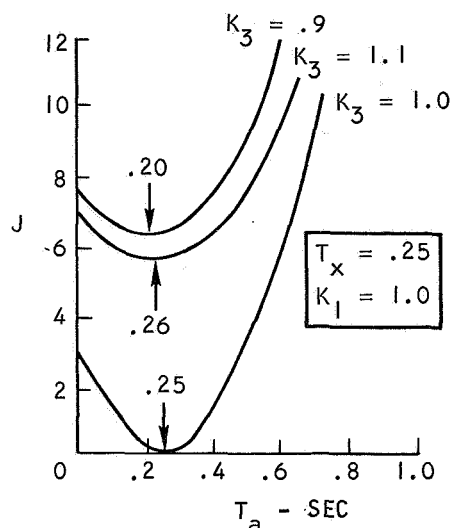


Figure 4.

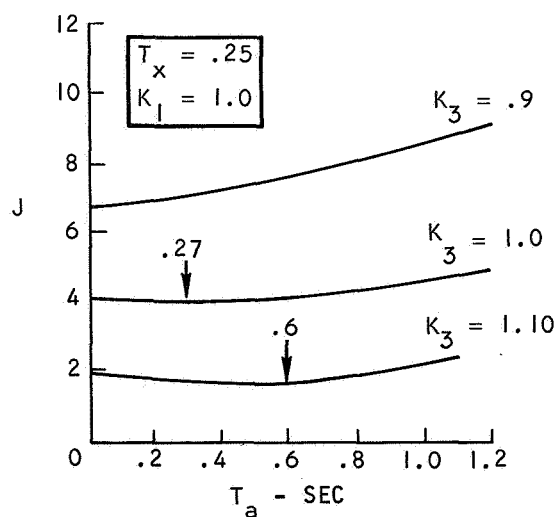


Figure 5.

TABLE 1.—SYSTEM AND MODEL FOR PROGRAMMED SEARCH FOR T_x .

EXPERIMENTS		SYSTEM	MODEL	FIG
1	FIRST ORDER	K_1/s	K_3/s	3
2	SECOND ORDER	$e^{-p_1 s} \frac{k_1(s+2)}{s(s+1)}$	$e^{-p_2 s} \frac{k_3(s+2)}{s(s+1)}$	4
3	—————	$K_1 \frac{(s+2)}{s(s+1)}$	$\frac{K_3}{s}$	5

CONCLUSION

Both gradient and programed search have been applied successfully to the identification of unknown sampling intervals in linear discrete-time systems. For those cases where the system form is known exactly and the sampling frequency is the only unknown, the identification rests on a solid analytical foundation. However, considerably more work is needed in the area of model-system mismatch before conclusions can be drawn.

REFERENCES

1. Bekey, G. A.; and Tomivic, R.: Sensitivity of Discrete Systems to Variation in the Sampling Interval. IEEE, PGAC, April 1966.
2. Freeman, H.: Discrete-Time Systems. John Wiley & Sons, Inc.

24. Cardinal Reconstruction Theory—A Tool for Estimating Effects of Display Scanning*

*Warren F. Clement
Systems Technology, Inc.*

A review of sampling processes and forms of continuous-signal reconstruction therefrom has revealed several forms which may be useful in modeling the average display scanning behavior of the human operator. The most promising form appears to be cardinal reconstruction. Low frequency approximations to the effective time delay for linear reconstruction are only slightly greater than for truncated cardinal reconstruction. The influence of sampled first derivative and of sample dwell time appear analogous to the influence of sample prediction in reducing the effective reconstruction time delay. Since these reconstruction processes have similar low-frequency responses, experimental attempts to isolate an effective reconstruction time delay in multidisplay-control tasks may be confounded by the reconstruction process of the skilled operator or pilot. Indeed, the skilled pilot may exhibit little or no incremental reconstruction time delay beyond the basic reaction time delay in a single task to which continuous attention is given.

INTRODUCTION AND MOTIVATION

Descriptions of continuous-signal reconstruction from sampling processes reveal properties which may help us to understand aircraft pilots' visual scanning behavior in the multidisplay-control context. Senders (ref. 1) has affirmed that predictions of the minimum frequency at which one among several displays must be observed is a sampling problem, and that existing sampled-data theory is adequate to apply to the modeling of experimental data intended for display design and evaluation purposes. Previous applications of periodic and aperiodic sampled-data stereotypes (refs. 2 to 10) have concentrated on internal models for the human operator in a single-display-control task. More recently, stochastic internal models have been suggested for describing the external visual sampling behavior of the human operator in a multidisplay-monitoring task (refs. 11 and 12). Here we shall apply some operational descriptions of continuous-signal reconstruction from sampled data and shall suggest implications of the results for pilots' effective scanning behavior in multidisplay-control tasks.

*This research was accomplished under Contract N00014-66-C0072 for the Office of Naval Research.

We begin with the implicit hypothesis affirmed by Senders (ref. 1) that visual scanning of several separate displays is evidence of multidisplay sampling by a pilot. A considerable body of experimental work (refs. 13 to 23) suggests that the observable eye fixation may be a virtually sufficient, but not necessary, manifestation of visual sampling. Exceptions to the hypothetical sufficiency condition may occur if no information is acquired on a particular fixation. For example, testimonial witnesses who look without seeing hardly satisfy sufficiency, but exceptions are difficult to prove among pilots under instrument flight rules. A model of the human operator with a foveal-peripheral perceptual switching policy has been proposed (ref. 24) to match recent measurements of operator-describing functions in a compensatory control task where fixation on the display was not necessary. Apparently, the single-task verbal-analytical describing-function models (e.g., ref. 25) provide a good basis for predicting the human controller's two-task behavior, especially where peripheral-visual perception is possible. However, Levison and Elkind (ref. 24) acknowledge that for display-control tasks of dimension greater than two, the complexity of the visual field may attenuate contributions from peripheral perception.

Under actual instrument-flight rules, a pilot's visual scanning policies are not perfectly periodic, yet they are not purely random. Following the work of ophthalmologists and optometrists prior to World War II, McGehee, Fitts, Senders, and others (refs. 11 to 23) have characterized pilots' visual scanning behavior in the following terms:

(1) Scan interval \bar{T}_s —the elapsed time between successive distinct fixations by the fovea on an assigned target; and

(2) Dwell time \bar{T}_d —the duration of a particular fixation.

Experimental results suggest that, for a particular and separate display among those required for instrument flight, these two measures approach independent quasi-random variation over observation time intervals between 1 and 4 minutes duration. (The average scan interval for the assembly of all displays will be more strongly correlated with the average duration of all fixations.) Scan interval is then analogous to sampling interval, and dwell time is analogous to the sample duration interval.

Scan interval and fixation frequency (cumulative number of distinct fixations in a unit of time) are reciprocally related. Hence, if an arithmetic mean be employed for one, a harmonic mean must be used for the other if the reciprocity is to be preserved between means, viz., $\bar{f}_s = 1/\bar{T}_s$, where \bar{f}_s is mean fixation frequency and $\bar{\omega}_s = 2\pi\bar{f}_s$. Formulation of hypotheses for average scan-interval prediction is preferably approached in the frequency domain (cf. ref. 26; hence, fixation frequency) rather than the time domain, because a theory of information sampling and reconstruction has already been formulated in the frequency domain (refs. 27 to 31).

While there are several sets of data (e.g., ref. 12) which lend plausibility to the presumption that fixation duration is a function of permissible error, input amplitude, and other quantities, a fixed average duration is a fair first approximation (cf. ref. 26).

Experimental work by Vossius and Wezel (refs. 32 and 33) with single-channel forced periodic visual sampling has produced manual tracking results in conformity with the Nyquist sampling theorem (refs. 27 and 28), which is the antecedent of the generalized sampling theorem (ref. 30).¹ The Nyquist theorem expresses a lower bound for sampling frequency (in terms of signal bandwidth) when no signal derivatives are sampled for use in reconstruction. Wezel employed sample duration intervals in the range 4 to 40 msec, whereas pilots' average fixation dwell times observed by McGehee, Fitts, et al. are in the range 500 to 800 msec where some form of signal reconstruction is probable and seldom less

¹The general sampling theorem, given magnitude plus R derivatives at each sample, is sampling rate $f_s \geq 2f_b/(R+1)$, where f_b is the effective bandwidth of the signal to be sampled.

than 400 msec even for monitoring a signal. If the pilot population employs increasing dwell time for derivative recognition, as Poulton's and Senders' results (refs. 34, 35, and 36) suggest, Wezel may have excluded derivative recognition in his experiment by his short sample exposure times; therefore, he is probably justified in seeking confirmation of the Nyquist theorem in contradistinction to its subsequent generalization.

Platzer and Krendel, reference 37, report tracking-error (but not describing-function) results in which first-derivative perception was forcibly excluded from a sampled display by employing a zero-order intersample hold. The controlled element was K/s^2 . Much larger average errors and many lost control divergences were observed when the first derivative was absent as compared with the same cases where it could be derived visually by the operator.

Thus we are clearly motivated to seek a theoretical relationship among at least sampling interval (or frequency), dwell time, and derivative recognition which will forecast an incremental effective scanning time delay which is useful for multidisplay-control task analysis. The theory of signal reconstruction from sampling data offers a basis for the search.

SAMPLING AT RANDOM

Bergen, reference 38, has examined an impulsive sampler followed by a pure gain. A periodic impulsive sampler acts like an impulsive modulator whose effective gain is inversely proportional to its sampling interval (refs. 39 and 40). In Bergen's example, the sampling interval of the modulator varies purely randomly, and the cascaded gain is equal to the mean sampling interval \bar{T}_s so that the modulator and gain have an effective input-to-output transmission of unity. It is shown theoretically that the output power spectral density of such a sampler is equivalent to that of a continuous-signal path with additive white noise whose power spectral density $\Phi_{nn}(\omega)$ is proportional to the product of mean sampling interval \bar{T}_s and mean square value e^2 of the signal to be sampled $e(t)$:

$$\Phi_{nn}(\omega) = \bar{T}_s \overline{e^2} \quad (1)$$

The additive white noise is uncorrelated with the signal input to the modulator. This result may have relevance for at least a portion of the remnant of a human operator in a multidisplay-control task. Biddle et al. (ref. 10) have observed dependence of broadband remnant on mean and variance of gaussian sampling interval distributions in a single simulated compensatory control task with forced random sampling of the error. Bergen's result clearly establishes mathematical justification for continuous analysis of an open-loop random impulsive sampling process and expresses the resulting average magnitude distortion (remnant) due to random variation in sampling interval \bar{T}_s . Bergen's result is also in consonance with macroscopic observations of pilots' virtually continuous control output—even in a multidisplay-control context with visual sampling. Naturally, any attendant phase distortion is lost in Bergen's average. However, any phase distortion which accompanies subsequent continuous signal reconstruction from sampled data is essential to the application of Bergen's model in a closed-loop multidisplay-control sampled task.

The present development will characterize reconstruction processes in the frequency domain for use with human-operator quasi-linear describing functions. We shall seek to supply the missing phase distortion which accompanies continuous-signal reconstruction by applying results from periodic sampling and reconstruction theory in terms of average properties for multidisplayed signal scanning.

PERIODIC IMPULSIVE SAMPLING AND LINEAR RECONSTRUCTION

Classical linear-signal reconstruction between samples of a function is based only on magnitude samples (refs. 39 and 40). Linear theory reconstructs estimates of signal derivatives by historical extrapolation exclusively from magnitudes. The increasing staleness of historical sampled magnitudes leads to extraordinarily great reconstruction time delay (and phase lag) incommensurate with observed pilot closed-loop behavior (refs. 40 and 41). Thus we shall reject historical sampled magnitude extrapolation alone as a model suitable for explaining pilots' signal reconstruction. Instead, we shall examine the frequency response of a linear-reconstruction process based on simultaneous impulsive samples of a function of time and its first-time derivative. Such a reconstruction function was employed in simulation by Bekey (ref. 7) and is expressible as

$$r(nT + \tau) = r(nT) + \lambda \tau r_1(nT) \quad (2)$$

where $t=nT$ are the impulsive sampling instants in time.

τ is the time of extrapolation (reconstruction) before the next sampling instant.

T is the (periodic) sampling interval which we shall identify with \overline{T}_S , the mean sampling interval for our quasi-random sampling processes.

$r(nT)$ is the sampled magnitude, whose reconstruction function is called a zero-order magnitude hold (ref. 40).

$r_1(nT)$ is the sampled first-time derivative, whose reconstruction function is also a zero-order (first-derivative) hold.

λ is a derivative weighting coefficient; $0 \leq \lambda \leq 1$.

The unit impulsive response $y_{h0}(\tau)$ corresponding to $r(nT)$ is

$$y_{h0}(\tau) = u(\tau) - u(\tau - T) \quad (3)$$

where $u(t)$ is the unit step function. The unit impulsive response $y_{h1}(\tau)$ corresponding to $\tau r_1(nT)$ is

$$y_{h1}(\tau) = \tau y_{h0}(\tau) \quad (4)$$

Figure 1 illustrates the linear combination of magnitude and first-derivative impulse responses to form an equivalent impulse response, each component of which is weighted respectively by the sampled function $f(nT)$ and its sampled derivative $f'(nT)$ to form $r(nT+\tau)$ in equation (2).

It can be shown by Laplace transformation and subsequent specialization in the frequency domain (ref. 42) that a low-frequency approximation for the overall sampling-and-reconstruction frequency response $Y_{ph}(j\omega)$ is

$$Y_{ph}(j\omega) \doteq e^{-j(1-\lambda)(\omega T/2)} ; \quad \frac{\omega T}{2} \ll 1 \quad (5)$$

Equation (5) shows that, for low-frequency signals, the phase lag of the effective reconstruction time delay in $Y_{ph}(j\omega)$ can be reduced as increasing weighting λ is given to first-derivative recognition. In the limit, as the derivative weighting coefficient λ approaches unity, the effective low-frequency reconstruction time delay will vanish, viz.,

$$\lim_{\lambda \rightarrow 1} Y_{ph}(j\omega) = 1 ; \quad \frac{\omega T}{2} \ll 1 \quad (6)$$

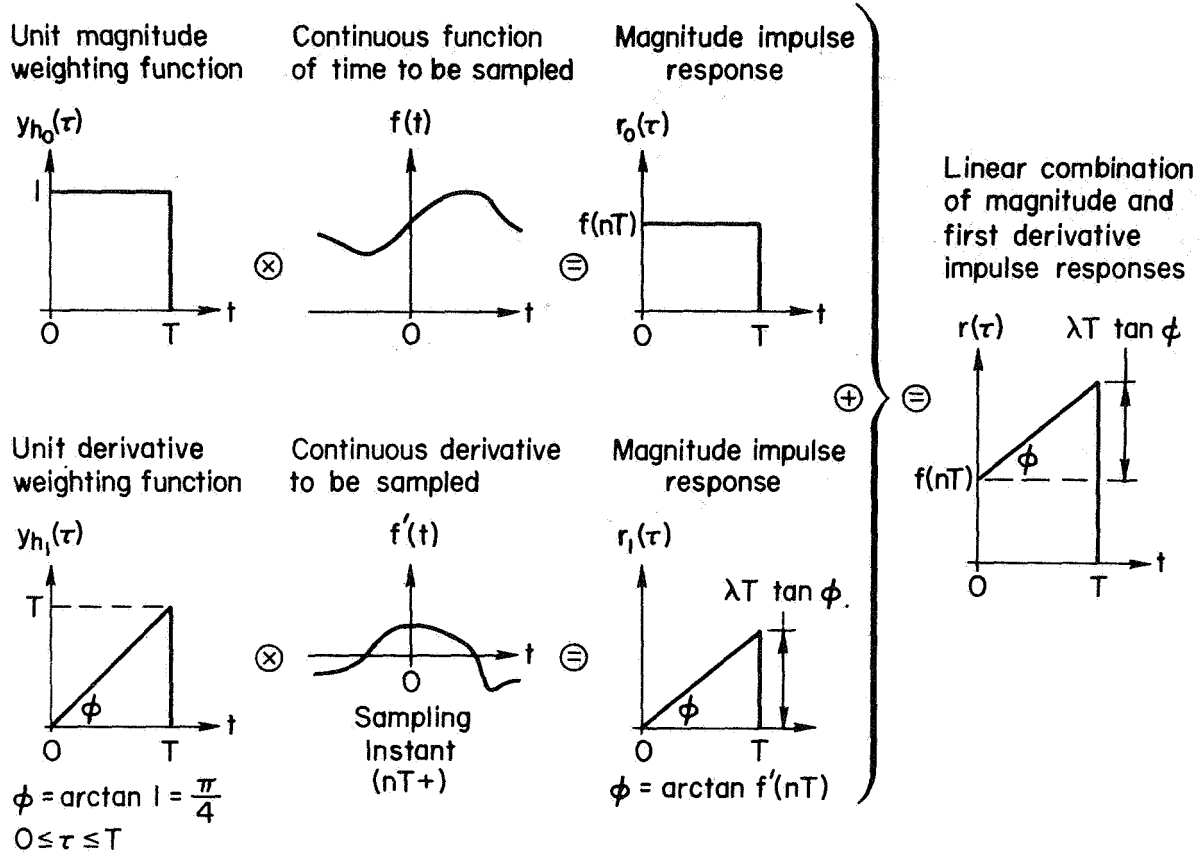


Figure 1. —Illustrating the linear combination of magnitude and first derivative impulse responses weighted, respectively, by the sampled function and its sampled derivative.

Therefore, even with linear reconstruction, low-frequency characteristics can approach the ideal limit of fidelity if the first derivative is sampled simultaneously with the signal magnitude and the members of the sample pair are weighted equally in the reconstruction algorithm.

LINEAR RECONSTRUCTION FOLLOWING A FINITE SAMPLE DWELL INTERVAL

We have agreed that one purpose of fixation dwell time may be to achieve near-simultaneity in signal magnitude and (at least) first-derivative samples. Alternatively, one may view the fixation dwell as an interval during which continuous control output occurs in response to the visual stimulus. These two views practically coalesce when one realizes that some form of intersample reconstruction may be common to both.

Farmanfarma (refs. 39 and 43) has developed a special case of the Laplace transform for operational analysis of finite-pulse-width sampled-data systems. It is of interest to apply an impulsive linear reconstruction process to Farmanfarma's notions. We shall apply a zero-order magnitude hold as an example, although the results of the previous section apply as well to simultaneous impulsive magnitude and first-derivative extrapolation following the finite-sample dwell interval.

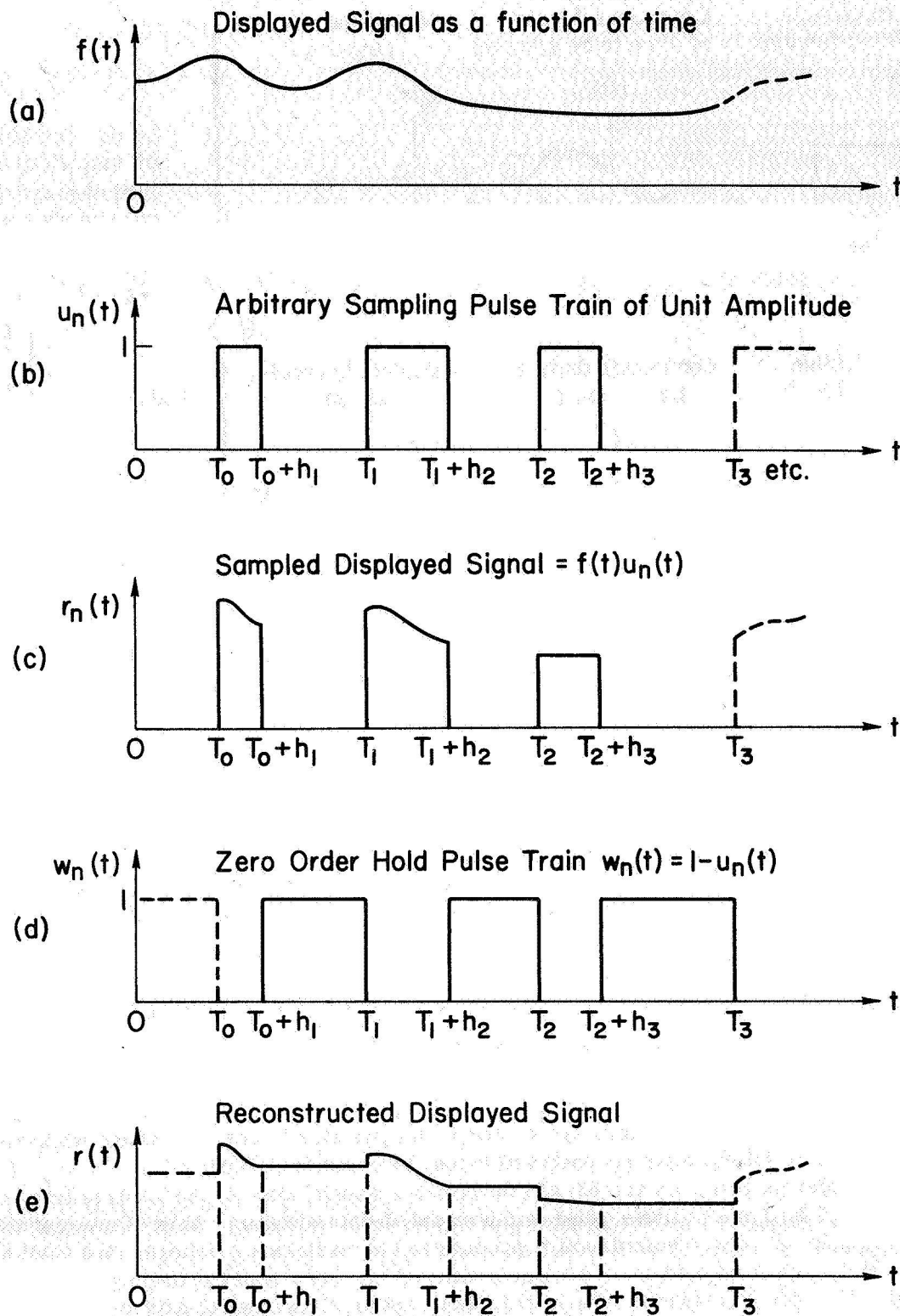


Figure 2.—Display perception model with finite pulse sampling and zero-order intersample reconstruction.

Figures 2(a), (b), and (c) illustrate the application of Farmanfarma's approach to a display-perception model with finite-pulse-width sampling where h_n is the dwell time. Figure 2(d) describes a zero-order hold pulse train and figure 2(e) shows the resulting reconstructed signal in combination with the finite-pulse samples. Our purpose will be to describe the effect of the zero-order reconstruction in the restricted case where sampling is synchronous ($T_0=0$) and periodic ($T_n=T$ and $h_n=h$; $n=1, 2, 3, \dots$). The restriction on dwell time insures synchronous reconstruction.

In the general case depicted by figure 2(d), the zero-order reconstruction unit pulse train is

$$w_n(t) = 1 - u_n(t) \quad (7)$$

where

$$u_n(t) = \begin{cases} 1, & T_n < t < T_n + h_{n+1}, \\ 0, & \text{otherwise} \end{cases} \quad n = 0, 1, 2, 3, \dots$$

In the special synchronous and periodic case,

$$w_p(t) = 1 - u_p(t) \quad (8)$$

where

$$u_p(t) = \begin{cases} 1, & nT < t < nT + h, \\ 0, & \text{otherwise} \end{cases} \quad n = 1, 2, 3, \dots$$

It can be shown (ref. 42) that the Laplace transform of the synchronous periodic zero-order reconstruction unit pulse train is composed of the product of two factors: $W_1(s)$, the transform of one zero-order reconstruction unit pulse; and $\frac{1}{1-e^{-Ts}}$, the transform of an infinite sum of identical periodic pulses.

The first reconstruction pulse (as well as all subsequent ones by the assumption of synchronism) is delayed by the dwell time interval h because of the arbitrary selection of temporal origin at the start of the finite sample. Therefore, we shall advance $w_1(t)$ by the dwell time interval to avoid any reconstruction penalty for delay in the zero-order hold. We have simply transferred our temporal origin to the start of reconstruction (fig. 3) in order to isolate contributions from the transfer function of only the reconstruction process $W_{1\Delta}(s)$ where $W_{1\Delta}(s)$ denotes the advanced transform of $W_1(s)$:

$$W_{1\Delta}(s) = e^{hs} W_1(s) = \frac{1 - e^{-(T-h)s}}{s} = \frac{1 - e^{-T(1-\eta)s}}{s} \quad (9)$$

where the dwell fraction $\eta = h/T$, $0 < \eta < 1$, and we shall identify

$$h \Rightarrow \bar{T}_d, \quad \text{mean dwell time}$$

$$T \Rightarrow \bar{T}_s, \quad \text{mean sampling interval}$$

Equation (9) can alternatively be written by inspection of figure 3(b), which shows the unit impulse response of the zero-order hold.

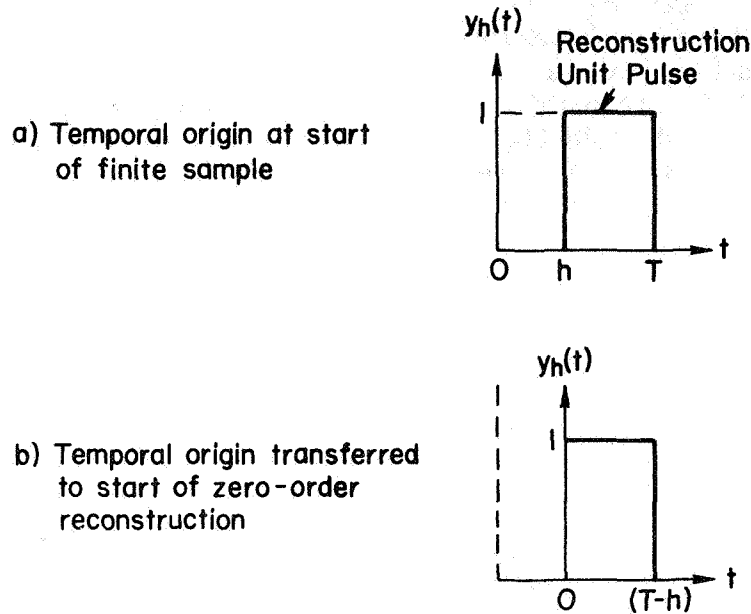


Figure 3.—Illustrating transfer of temporal origin from start of finite sample to start of reconstruction.

The overall sampling-and-reconstruction frequency response ($s \rightarrow j\omega$) is

$$Y_{Ph}(j\omega) = \frac{1}{T} W_{1\Delta}(j\omega) = (1-\eta) \frac{[\sin \omega T(1-\eta)/2]}{[\omega T(1-\eta)/2]} e^{[-j\omega T/2](1-\eta)} \quad (10)$$

The low-frequency approximation with $T \Rightarrow \bar{T}_s$,

$$Y_{Ph}(j\omega) \doteq s(1-\eta)e^{[-j\omega \bar{T}_s/2](1-\eta)} ; \quad [\omega \bar{T}_s/2](1-\eta) \ll 1 \quad (11)$$

can be obtained by simplification of equation (10) with a first-order approximation for the sine. Now by comparing equations (5) and (11), we notice that, for low-frequency signals, the phase lag of the effective reconstruction time delay in $Y_{Ph}(j\omega)$ (eq. (11)) can be reduced by increasing the dwell fraction ($\eta = h/T \Rightarrow \bar{T}_d/\bar{T}_s$) in a manner analogous to the reduction achieved in equation (5) by first-derivative recognition. In the limit as dwell time approaches the sampling interval, the sampled process becomes a continuous transmission process, and the reconstruction process vanishes, viz.,

$$\lim_{\eta \rightarrow 1} Y_{Ph}(j\omega) \doteq 0 ; \quad \frac{\omega \bar{T}_s}{2} \ll 1 \quad (12)$$

Therefore, low-frequency phase characteristics are similar for both points of view of the linear reconstruction processes: (1) with simultaneous first derivative and magnitude recognition from impulsive samples, or (2) with an impulsive zero-order hold following finite-pulse samples.

TRUNCATED CARDINAL RECONSTRUCTION

Cardinal-reconstruction theory is an outgrowth of Nyquist's early sampling reconstruction studies (ref. 27) eventually proved by Shannon (ref. 28) and recently generalized to include derivative sampling by Fogel, Linden, and Abramson (refs. 29, 30, and 31). This theory treats the perfect reconstruction of a frequency band-limited signal from only the essential "cardinal" samples prescribed by the lower frequency bound of the generalized (or Nyquist) sampling theorem (or, conversely, the upper bound of the Nyquist interval). Cardinal reconstruction is exact only in an infinite length of time before and after each impulsive sample. Therefore, cardinal reconstruction demands prediction or preview to achieve perfect reconstruction in time from the assembly of sample weighting functions. The capability of the theory to treat prediction may be an advantage when we apply the notion of cardinal reconstruction to the human pilot, since Vossius and Wezel (refs. 32 and 33) report evidence of precognition. A model which will describe mathematically consistent reversible transitions between derivation and prediction is desirable.

The reconstruction transfer function Y_{ph} will have an impulse response identical to the sample weighting function in a finite time interval equal to the intersample interval with the freedom of derivation from the sample or prediction of the sample or some combination thereof. We shall call this operation truncated cardinal reconstruction, since we shall use the same (cardinal) family of weighting functions with finite intersample truncation. Then we shall derive the desired phase (and amplitude) distortion characteristics in the complex operational domain (which can be specialized for the frequency domain by $s \rightarrow j\omega$) for the reconstruction transfer function Y_{ph} by (bilateral) Laplace transformation of the truncated-cardinal-impulse response.

ZERO-ORDER RECONSTRUCTION.—Figure 4 illustrates the response of zero-order cardinal reconstruction from an impulsive sample. (Zero-order in this case means without recognition of signal derivatives.) For this case, the (Nyquist) sampling theorem prescribes the lower frequency bound $2\pi/T = 2\omega_b$ where ω_b is the low-pass (rectangular) frequency bandwidth of the signal to be sampled. Therefore, $T = \pi/\omega_b$ as illustrated by the dual time scale

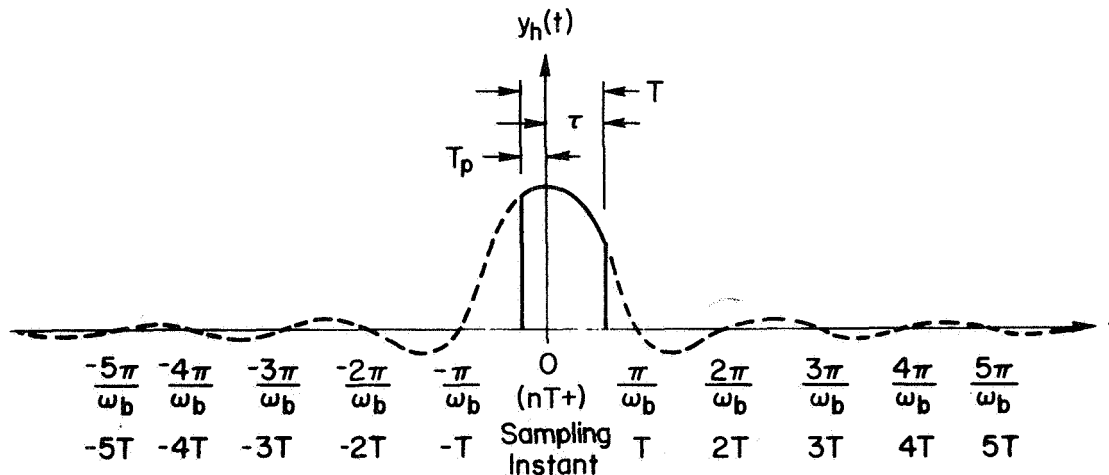


Figure 4. —Truncated impulse response of zero-order cardinal reconstruction filter.

in figure 4. We shall confine truncation to intervals not exceeding one sampling interval and shall identify a prediction time interval T_p as that interval by which the start of the truncated impulse response anticipates the sampling instant. Then,

$$T_p = \tau - T \leq 0 \quad (13)$$

where τ is a truncation interval following the instant of sampling.

The transfer function of the (truncated) zero-order cardinal reconstruction filter is the Laplace transform of the corresponding truncated impulse response, where it should be understood that the bilateral Laplace transform must be used if $-T < T_p < 0$. Since the truncated impulse response is

$$y_{ph}(t) \begin{cases} = \frac{\omega_b}{\pi} \left(\frac{\sin \omega_b t}{\omega_b t} \right) & ; \quad T_p \leq t \leq \tau \\ = 0 & ; \quad T_p > t > \tau \end{cases} \quad (14)$$

and the truncated unilateral Laplace transform ($0 \leq t$) is

$$Y_{ph}(s) = \frac{\omega_b}{\pi} \int_0^{\tau} e^{-st} \frac{\sin \omega_b t}{\omega_b t} dt ; \quad \begin{cases} 0 \leq \tau \leq T \\ 0 \leq s \end{cases} \quad (15)$$

or with obvious interchange of limits such that $-T \leq T_p \leq 0$, if $t \leq 0$ and $s \leq 0$. Since $y_{ph}(t)$ is even, the bilateral Laplace transform in the symmetrically truncated case $|\tau| = |T_p| = T/2$ will be twice the unilateral transform.

The transform integral equation (15) is not readily evaluated, except in the classical case of an infinite upper limit (without truncation). However, the truncated transform equation (15) can be integrated by parts under restricted conditions which are compatible with the bandwidth limitation imposed by the sampling theorem (ref. 42). The frequency response of sampling and reconstruction with $T \Rightarrow \bar{T}_s$ is given approximately by

$$Y_{ph}(j\omega) \doteq \frac{\sqrt{2}}{\pi} e^{-j\omega(4\bar{T}_s/3\pi)} ; \quad 0 \leq t \leq \bar{T}_s, \quad 0 \leq \frac{\omega}{\omega_b} \leq 1 \quad (16a)$$

and

$$Y_{ph}(j\omega) \doteq \frac{3\sqrt{3}}{2\pi} e^{-j\omega(2\bar{T}_s/3\pi)} ; \quad -\frac{\bar{T}_s}{2} \leq t \leq \frac{\bar{T}_s}{2}, \quad 0 \leq \frac{\omega}{\omega_b} \leq 1 \quad (16b)$$

Thus, the effective time delay for zero-order truncated cardinal reconstruction after the sampling instant (eq. (16a)) is valid over the reconstruction frequency bandwidth, yet it is slightly less than that for zero-order linear construction at low frequencies. (Cf. eq. (5) for $\lambda = 0$.) If the (truncated) cardinal-impulse response anticipates the sampling instant by one-half the period, then the effective time delay for reconstruction (with partial preview or prediction) is halved over that without preview or prediction. The influence of sample

prediction here appears analogous to the influence of derivative weighting coefficient λ in equation (5) and the influence of dwell time $h \Rightarrow \bar{T}_d$ in equation (11). Since these reconstruction processes have similar operational descriptors, experimental attempts to isolate an effective reconstruction time delay in multidisplay-sampling contexts may be confounded by the reconstruction process of the skilled operator or pilot. In practice, the reconstruction time delay may not even be discernible in measurements, unless fixation dwell, first-derivative recognition and preview or precognition by the subject are deliberately excluded. The incremental reconstruction time delay, then, would appear to be one inverse measure of the skill of a subject in a multidisplay-control task.

FIRST-ORDER RECONSTRUCTION.—Linden (ref. 31) presents the complete impulse response for first-order cardinal reconstruction from simultaneous (periodic) impulsive samples of signal magnitude and first derivative. The generalized sampling theorem in this case prescribes the lower frequency bound $2\pi/T = \omega_b$; therefore $T = 2\pi/\omega_b$ as illustrated by the dual time scale in figure 5, which defines truncation of the magnitude and first-derivative impulse responses given, respectively, by

$$y_{ph0}(t) \begin{cases} = \frac{\omega_b}{2\pi} \left[\frac{\sin(\omega_b t/2)}{(\omega_b t/2)} \right]^2 & ; T_p \leq t \leq \tau \\ = 0 & ; T_p > t > \tau \end{cases} \quad (17)$$

and
$$y_{ph1}(t) = t y_{ph0}(t) \quad (18)$$

The Laplace transforms of the truncated magnitude and first-derivative impulse response can be derived by a procedure analogous to that for zero-order truncated reconstruction (ref. 42). The frequency response of sampling and reconstruction with $T \Rightarrow \bar{T}_s$ is given approximately by

$$Y_{ph}(j\omega) \doteq \frac{\sqrt{3}}{\pi} e^{[-j(1-\lambda)\omega(4\bar{T}_s/3\pi)]} \quad ; \quad 0 \leq t \leq \bar{T}_s \quad (19)$$

and
$$Y_{ph}(j\omega) \doteq \frac{3\sqrt{3}}{2\pi} e^{[-j(1-\lambda)\omega(2\bar{T}_s/3\pi)]} \quad ; \quad -\frac{\bar{T}_s}{2} \leq t \leq \frac{\bar{T}_s}{2}$$

where $0 \leq \omega/\omega_b \leq 1$ and $0 \leq \lambda \leq 1$ is derivative weighting coefficient. This effective time delay for first-order truncated cardinal reconstruction is compatible with the frequency bandwidth limitation imposed by the sampling theorem and is also nearly the same as that found for the corresponding linear reconstruction (eq. (5)) at low frequencies. In addition, the predictive degree of freedom in the cardinal model provides a consistent descriptor for including effects of preview or precognition on reconstruction.

SUMMARY AND CONCLUSIONS

Three forms of signal reconstruction from sampling processes have been investigated for their usefulness in modeling average sampling-and-reconstruction behavior of the human operator:

- (1) Linear reconstruction from impulsive first-derivative samples recognized simultaneously with impulsive magnitude samples
- (2) Linear reconstruction following finite pulse-width samples
- (3) Truncated cardinal reconstruction

The most significantly useful classification appears to be truncated cardinal reconstruction because (1) it is related to the frequency bandwidth of the signal to be sampled by the generalized (Nyquist) sampling theorem; (2) it can accommodate derivative recognition, the low-frequency-phase-angle contributions of which are indistinguishable from those of reconstruction following finite pulse width samples; and (3) it can accommodate preview, prediction, or precognition.

The transfer function (and frequency response) of truncated cardinal reconstruction is that of a pure gain and time delay valid over the frequency bandwidth of the signal from which samples are taken in conformity with the generalized sampling theorem.

A summary of time-delay increments is given in table 1 for the two classes of reconstruction investigated following a finite sample dwell of average interval \bar{T}_d . Although based on reconstruction from periodically impulsively sampled data, the two classes are identified with mean values from distributions of varying sampling interval and dwell interval, based on the continuous equivalent model of Bergen (ref. 38) for a randomly varying sampling interval. A graphic presentation of the relationship between fixation frequency (scanning or sampling frequency) and incremental time delay is given in figure 6 for first-order truncated cardinal reconstruction following fixation dwell intervals which are typical on instrument

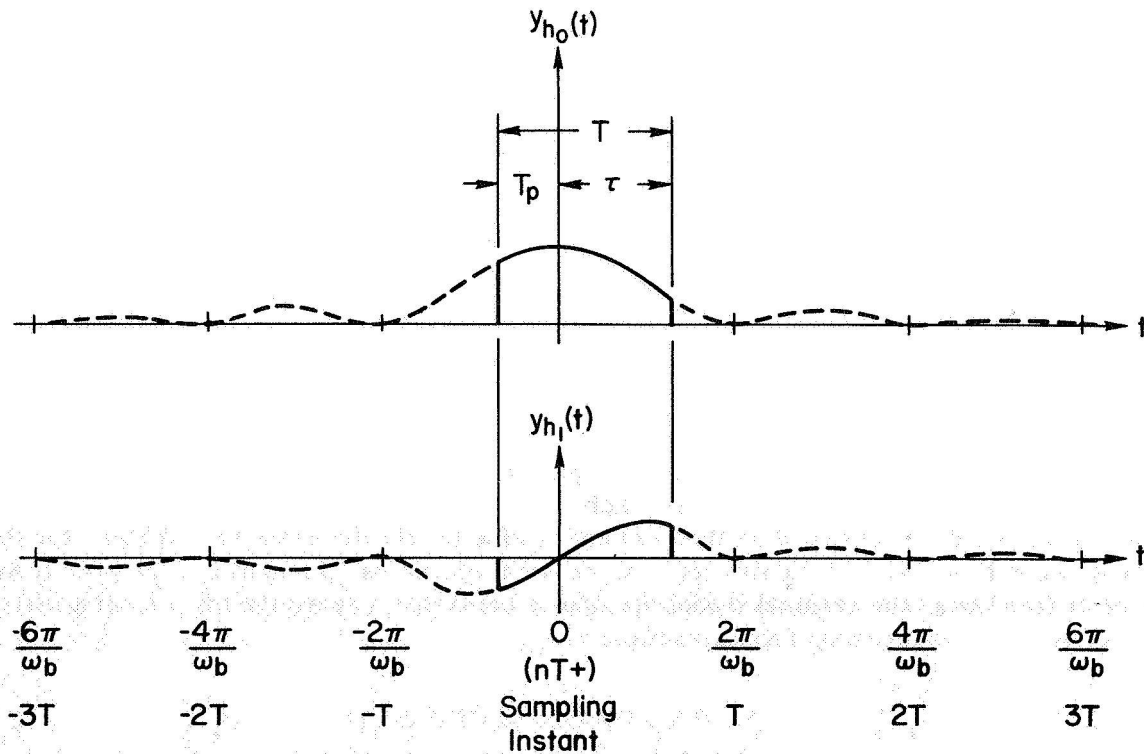


Figure 5. —Truncated magnitude and first derivative impulse responses for first-order cardinal reconstruction filters.

TABLE 1.—EFFECTIVE RECONSTRUCTION TIME DELAY INCREMENT
ASSOCIATED WITH SAMPLING

RECONSTRUCTION CLASS	HIGHEST SIGNAL DERIVATIVE EMPLOYED	TIME DELAY INCREMENT $\Delta\tau$
Linear	None	$\frac{\bar{T}_s}{2} \left(1 - \frac{\bar{T}_d}{\bar{T}_s} \right)$
	First (with weighting coefficient λ ; $0 \leq \lambda \leq 1$)	$(1-\lambda) \frac{\bar{T}_s}{2} \left(1 - \frac{\bar{T}_d}{\bar{T}_s} \right)$
Truncated-Cardinal	None	$\frac{4\bar{T}_s}{3\pi} \left(1 - \frac{\bar{T}_d}{\bar{T}_s} \right)$
	First (with weighting coefficient λ ; $0 \leq \lambda \leq 1$)	$(1-\lambda) \frac{4\bar{T}_s}{3\pi} \left(1 - \frac{\bar{T}_d}{\bar{T}_s} \right)$

\bar{T}_s mean sampling interval

\bar{T}_d mean dwell time

λ first derivative weighting coefficient

approaches. Derivative recognition is influential in reducing the reconstruction time delay and in reducing its sensitivity to changes in scanning frequency.

The low-frequency effective time delay for linear reconstruction is only slightly greater than that for truncated cardinal reconstruction. In fact the two forms are identical when first-order Pade' approximants are employed. If the (truncated) cardinal-impulse response anticipates the sampling instant by one-half the period, then the effective time delay for reconstruction (with partial preview or prediction) is halved over that without preview or prediction.

The influence of sample prediction appears analogous to the influence of derivative weighting coefficient and the influence of dwell time. Since these reconstruction processes exhibit similarity among their operational descriptors, experimental attempts to isolate an effective reconstruction time delay in multidisplay-sampling and control tasks may be confounded by the reconstruction process of the skilled operator or pilot. The incremental reconstruction time delay, then, would appear to be one inverse measure of the skill of a subject in a multidisplay-control context. Indeed, the pilot skilled at signal reconstruction from several sampled displays may exhibit little or no incremental reconstruction time delay beyond that basic reaction time delay in a single task to which continuous attention is given.

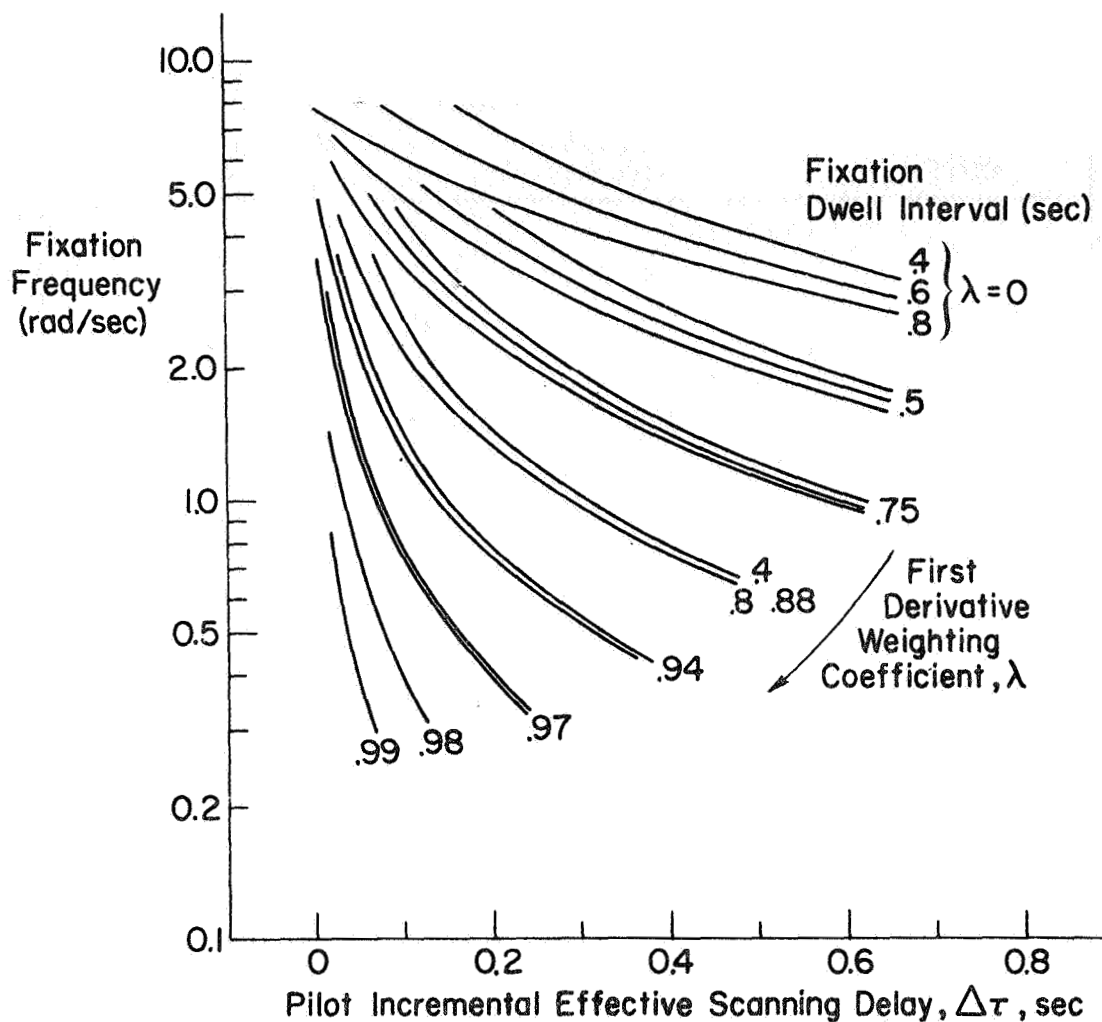


Figure 6.—Theoretical fixation frequency, dwell interval and effective scanning delay relationship.

REFERENCES

1. Senders, J. W.: Man's Capacity to Use Information From Complex Displays. Information Theory in Psychology, H. Quastler, ed., The Free Press, Glencoe, Ill., 1955.
2. Craik, J.: Theory of the Human Operator in Control Systems; I, The Operator as an Engineering System; II, Man as an Element in a Control System. Brit. J. Psychol., Dec. 1947 and Mar. 1948.
3. North, J. D.: The Human Transfer Function in Servo Systems. Directorate of Weapons Research, (British) Ministry of Defense, Rept. No. WRD/6/50, 1950. (Reprinted in Automatic and Manual Control, ed. by A. Tustin, Butterworths, London, 1958, pp. 473-502.)

4. Hayes: Characteristics of the Human Operator as an Element in a Closed-Loop Control System. Directorate of Weapons Research, (British) Ministry of Defense, Rept. No. WRD/9/53, 1953.
5. Hayes: The Intermittent Nature of the Response of a Human Operator. Royal Military College of Science, Electrical Engineering Branch, Rept. 1/54, 1954.
6. Ward, J. R.: The Dynamics of a Human Operator in a Control System; A Study Based on the Hypothesis of Intermittency. University of Sydney, Dept. of Aeronautics, Ph.D. Thesis, May 1958.
7. Bekey, G. A.: Sampled Data Models of the Human Operator in a Control System. UCLA, Dept. of Engineering, Ph.D. Thesis, Jan. 1962. (Reprinted as ASD-TR-62-36, Feb. 1962.)
8. Bekey, G. A.: The Human Operator as a Sampled-Data System. IRE Trans., vol. HFE-3, no. 2, Sept. 1962, pp. 43-51.
9. Lemay, L. P.; and Wescott, J. H.: The Simulation of Human Operator Tracking Using an Intermittent Model. Paper presented to the IRE Intern. Cong. Human Factors Electron., Long Beach, Calif., 3-4 May 1962.
10. Biddle, J.; Jacobsen, A.; and Bekey, G. A.: The Effect of Random Sampling Intervals on Sampled Data Models of the Human Operator. Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967, pp. 247-258.
11. Senders, J. W.: The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. IEEE Trans., vol. HFE-5, Sept. 1964, pp. 2-5.
12. Senders, J. W.; Elkind, J. I.; Grignetti, M. C.; and Smallwood, R. D.: An Investigation of the Visual Sampling Behavior of Human Observers. NASA CR-434, Apr. 1966.
13. Cole, E. L.; Milton, J. L.; and McIntosh, B. B.: Routine Maneuvers Under Day and Night Conditions, Using an Experimental Panel Arrangement. WADC-TR-53-220, Mar. 1954.
14. Jones, R. E.; Milton, J. L.; and Fitts, P. M.: Eye Fixations of Aircraft Pilots; I. A Review of Prior Eye Movements Studies and a Description of a Technique for Recording the Frequency, Duration, and Sequences of Eye Fixations During Instrument Flight. Air Materiel Command AF TR-5837, Sept. 1949 (includes summary of McGehee's studies circa 1944).
15. Fitts, P. M.; Jones, R. E.; and Milton, J. L.: Eye Fixations of Aircraft Pilots; III. Frequency, Duration, and Sequence of Fixations when Flying Air Force Ground Controlled Approach System (GCA). Air Materiel Command AF TR-5967, Feb. 1950.
16. Jones, R. E.; Milton, J. L.; and Fitts, P. M.: Eye Fixations of Aircraft Pilots; IV. Frequency, Duration, and Sequence of Fixations During Routine Instrument Flight. Air Materiel Command AF TR-5975, Dec. 1949.
17. Fitts, P. M.; Jones, R. E.; and Milton, J. L.: Eye Movements of Aircraft Pilots During Instrument-Landing Approaches. Aeron. Eng. Rev. vol. 9, no. 2, Feb. 1950, pp. 24-29.
18. Milton, J. L.; Jones R. E.; and Fitts, P. M.: Eye Fixations of Aircraft Pilots; II. Frequency, Duration, and Sequence of Fixations When Flying the USAF Instrument Low Approach System (ILAS). Air Materiel Command AF TR-5839, Oct. 1949.
19. Milton, J. L.; McIntosh, B. B.; and Cole, E. L.: Eye Fixations of Aircraft Pilots; VI. Fixations During Day and Night ILAS Approaches Using an Experimental Instrument Panel Arrangement. Air Materiel Command AF TR-6570, Oct. 1951.

20. Milton, J. L.; and Wolfe, F. J.: Fixations During Zero-Reader Approaches in a Jet Aircraft. WADC-TR-52-17, Feb. 1952.
21. Watts, A. F. A.; and Wiltshire, H. C.: Investigation of Eye Movements of an Aircraft Pilot Under Blind Approach Conditions. The College of Aeronautics, Note no. 26, May 1955.
22. Lennox, D.: Airline Pilots' Eye Movements During Take-Off and Landing in Visual Meteorological Conditions. Aeronautical Research Labs, Human Engineering Note 15, Aug. 1963.
23. Winblade, R. L.: Current Research on Advanced Cockpit Display Systems. AGARD Rept. No. 491, Oct. 1964.
24. Levison, W. H.; and Elkind, J. I.: Studies of Multivariable Manual Control Systems; Two-Axis Compensatory Systems with Separated Displays and Controls. NASA CR-875, Oct. 1967.
25. McRuer, D. T.; Graham, D.; Krendel, E. S.; and Reisener, W., Jr.: Human Pilot Dynamics in Compensatory Systems. AFFDL-TR-65-15, July 1965.
26. Clement, W. F.; Graham, D.; and Best, J. J.: A Reexamination of Eye Movement Data. Systems Technology, Inc., TM 163-A, 28 Feb. 1967.
27. Nyquist, H.: Certain Topics in Telegraph Transmission Theory. Trans. AIEE, vol. 47, Apr. 1928, p. 617.
28. Shannon, C. E.: Communication in the Presence of Noise. Proc. IRE, vol. 37, no. 1, Jan. 1949, pp. 10-21.
29. Fogel, L. J.: A Note on the Sampling Theorem. IRE Trans. on Information Theory, Dec. 1956.
30. Linden, D. A.; and Abramson, N. M.: A Generalization of the Sampling Theorem. Information and Control, vol. 3, no. 1, 1960, pp. 26-31.
31. Linden, D. A.: A Discussion of Sampling Theorems. Proc. IRE, vol. 47, no. 7, July 1959, pp. 1219-1226.
32. Vossius, G.: Die Vorhersageeigenschaften des Systems der Willkurbewegung. Neuere Ergebnisse der Kybernetik, R. Oldenbourg, Munich, 1964. The Prediction Capabilities of the System of Haphazard Motion, Systems Technology, Inc., Tech. Translation no. 2, July 1965.
33. Wezel, F.: Untersuchungen Über die Willkurbewegung der Menschlichen Hand Mit Getastet Dargebotenen Reizmustern. Thesis, J. W. Goethe Univ., Frankfurt am Main, 1962. Investigations of Haphazard Movement of Human Hand With Sampled Stimuli, Systems Technology, Inc., Tech. Translation No. 3, July 1965.
34. Poulton, E. C.: Perceptual Anticipation in Tracking with Two-Pointer and One-Pointer Displays. Brit. J. Psychol., vol. 43, 1952, pp. 222-229.
35. Poulton, E. C.: The Basis of Perceptual Anticipation in Tracking. Brit. J. Psychol., vol. 43, 1952, pp. 295-302.
36. Senders, J. W.: Tracking With Intermittently Illuminated Displays. WADC TR 55-378, Oct. 1955.
37. Platzner, H. L.; and Krendel, E. S.: A Non-Linear Approach to Human Tracking. Franklin Institute Interim Tech. Rept. No. I-2490-1, 21 Dec. 1955.
38. Bergen, A. R.: On the Statistical Design of Linear Random Sampling Schemes. Proc. IFAC, vol. 1, Butterworth (London), 1961, pp. 430-436.
39. Jury, E. I.: Sampled-Data Control Systems. John Wiley & Sons, 1958.
40. Ragazzini, J. R.; and Franklin, G. F.: Sampled Data Control Systems. McGraw-Hill, 1958.

41. Magdaleno, R. E.: Sampling Behavior for the Human Operator. Systems Technology, Inc., TM 115-4, 27 July 1962.
42. Clement, W. F.: Cardinal Reconstruction Theory—A Tool for Estimating Effects of Display Scanning. Systems Technology, Inc., TM 163-B, 1 Mar. 1967.
43. Farmanfarma, G.: Analysis of Sampled-Data Systems With Finite Pulse Width, Ph.D. Thesis, Dept. of Electrical Engineering, Univ. of California, Berkeley, 1957.

VI. EFFECTS OF ENVIRONMENTAL VARIABLES ON PERFORMANCE

25. On Biocybernetics of the Vestibular System*

*Laurence R. Young
Massachusetts Institute of Technology*

This paper describes the characteristics of the human vestibular system modeled from the point of view of control theory. The vestibular control system in man is one of several systems used to determine his orientation in space. The three semicircular canals in each inner ear, oriented in roughly orthogonal planes, respond to angular acceleration about an axis normal to the plane of the canal. The otoliths are stimulated by linear acceleration as well as the gravity field (specific force).

Identification of the system output requires careful consideration. One may draw the analogy to inertial navigation, with the semicircular canals providing the attitude signals usually generated by gyroscopes, and the otoliths providing the linear acceleration signals to be twice integrated, yielding position. Orientation with respect to the apparent vertical forms an important output used for the postural reflex and for manual control of vehicles.

An additional output of the vestibular system is the set of control signals sent to the extraocular muscles. It is well known that the vestibular system, when stimulated, drives the eyes conjugately in the direction opposite to perceived rotation, thereby tending to stabilize the eyes in space despite rotations of the head. The importance of nystagmus as an indicator of vestibular function and otolith-canal cross-coupling is discussed in detail.

The familiar torsion-pendulum model of the horizontal canal has been remarkably successful in describing most of the subjective and objective responses; our experiments on thresholds for low angular accelerations provide further quantitative checks on this model. We show that the dynamics of the vertical canals are different from those of the horizontal canals and that these differences in dynamics in some way could explain the subjective feeling in an aircraft that the instantaneous axis of rotation is different from that of the objective instantaneous axis of rotation. The primary tool in this study was the measurement of latency of sensation to constant angular acceleration around a vertical axis as a function of acceleration level. Fitting the data with the exponential relation resulting from the overdamped accelerometer equation yields the long-time constant. The long-time constant for the sensation of rotation about the sagittal (roll) axis was approximately 7 seconds, compared to the 10 to 12 seconds found for the horizontal canals. The vertical canal threshold was found to be approximately 0.5 deg/sec^2 compared to approximately 0.14 deg/sec^2 for the

*This research was supported in part by NASA under Grant NsG-577. The important contributions of Prof. J. L. Meiry, Prof. Y. T. Li, and the author's graduate students in the Man-Vehicle Laboratory are gratefully acknowledged. This paper was originally prepared for presentation at the Edsel B. Ford Institute symposium, "Biocybernetics of the Central Nervous System," February 1968, and the IFAC symposium on "Technical and Biological Problems in Cybernetics," October 1968. The full text of the paper is available from the Man-Vehicle Laboratory, MIT, Cambridge, Mass. 02139.

horizontal plane. Nonlinearities include asymmetric thresholds, saturation, and sensitivity. A feedback loop indicating habituation or the decreasing sensitivity to repeated stimulation is included.

Control system descriptions of the gravireceptors had been almost nonexistent prior to our studies. Experiments were difficult to perform when they required stimulating the linear acceleration sensors without simultaneous stimulation of the semicircular canals, and consequently quantitative linear sensor data that would be useful in a dynamic orientation model were quite limited. To develop a pure linear acceleration as the stimulus, we constructed a 32-foot horizontal track on which a vehicle is driven under fine position control, providing a useful tool for studying response to low-level acceleration. The studies correlating human input-output data using such a device define the dynamic response of the linear-translation sensors and can only inferentially be applied to the otolith characteristics.

An important early experiment was to establish the phase relation between the subjective sensation of linear velocity and the objective linear velocity for sinusoidal linear oscillation. A linear model was developed on the basis of this phase relationship. The nonlinearities corresponding to otolith threshold were included and checked by means of latency to sensation of linear acceleration. Recent revision of the otolith model to agree with electrophysiological data is further supported by experiments on dynamic counterrolling of the eyes about the gaze axis. The complete vestibular model presented in this paper describes most nonvisual dynamic orientation phenomena in a manner familiar to control engineers. Practical samples are included to show the way vestibular stimulation contributes to the pilot's control compensation in a closed-loop vehicle control task.

26. Motion Cues in Man-Vehicle Control*

Richard S. Shirley
Electronics Research Center, NASA
Laurence R. Young
Massachusetts Institute of Technology

The human operator's use of roll-angular-motion cues in man-vehicle control is investigated. Extensive data for the human operator's describing function and remnant are taken for a wide range of vehicle dynamics under conditions of visual cues only, roll-motion cues only, and simultaneous visual and roll-motion cues. Addition of roll-motion cues to visual cues permits the human operator to increase his phase lead at frequencies above 2 rad/sec, thereby allowing higher gain and crossover frequency and reduced tracking error.

INTRODUCTION

If a man is to be in the control loop of a proposed vehicle, it is necessary to consider his control characteristics from the outset of the vehicle design. In the past, this problem has generally led to the simulation of the proposed vehicle dynamics on an analog computer and the testing of the pilot's ability to control the simulated dynamics in the laboratory. Unfortunately, the results of such fixed-base experiments did not always agree with later experiments performed with the actual aircraft. To obtain the correct results in the laboratory, additional simulations of the proposed vehicle's cab, sounds, and motions were made. It was learned that in many cases the simulation of vehicle angular motions was of primary importance. When the vehicle angular motions were simulated in the laboratory, laboratory data were often in close agreement with later flight data. Young (ref. 1) attributed this to the fact that angular-motion cues frequently are used to stabilize vehicle orientation prior to controlling vehicle position and, thus, become of primary importance. For example, before stabilization of the lateral position of a hovering helicopter, it is necessary to stabilize its roll angle.

Because motion simulation is difficult and expensive, as is the use of variable-stability aircraft, it would be useful to be able to predict actual in-flight data for the human operator from laboratory fixed-base experiments. The results presented in this paper, which concern the human operator's use of roll motion cues, represent a step in this direction.

*This paper is based on research performed at the Man-Vehicle Laboratory, MIT, under NASA Grant NsG-577.

THE EXPERIMENT

The following experiment was performed to obtain data for the human operator to determine how he uses roll-motion cues in vehicle control. Data were taken while the human operator was presented visual cues only, roll-motion cues only, and simultaneous visual and roll-motion cues. Data were also taken for a wide range of vehicle dynamics, and for a range of other experimental parameters. The vehicle dynamics tested included

$$Y_c(s) = \frac{K}{s(\tau s + 1)} \quad , \quad K \text{ and } \tau \text{ vary} \quad (1)$$

$$Y_c(s) = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \quad , \quad \omega_n^2 \text{ and } 2\xi\omega_n \text{ vary} \quad (2)$$

$$Y_c(s) = \frac{\omega_n^2}{s(s^2 + 2\xi\omega_n s + \omega_n^2)} \quad , \quad \omega_n^2 \text{ and } 2\xi\omega_n \text{ vary} \quad (3)$$

where s is the Laplace operator. The dynamics given by equations (1) through (3) are of general interest. For example, among other things equation (1) corresponds to simplified airplane lateral dynamics (for various values of K and τ), relating aileron angle to roll angle. Similarly, equation (3) corresponds to some compensated helicopter dynamics (with roll-angle feedback), relating control-stick position to horizontal position. The other experimental parameters which were varied (one at a time) include control-stick gain, system input power, and the input-spectrum breakpoint.¹

Subjects were placed in the cab of a simulator capable of producing roll motions, and the vehicle dynamics were simulated on an analog computer. The subjects' bodies were tightly strapped in place, but their heads were not directly restrained. The cab of the simulator was covered, and contained an oscilloscope and a control stick. The system attitude error was displayed to the man as a horizontal displacement of a dot from the centered position on the oscilloscope face (for visual cues), and as a rotation in roll of the simulator cab from the vertical (motion cues). The dynamics of the simulator were nearly equal to a pure delay of 0.1 second ($e^{-0.1s}$) over the frequency range of interest (0.1 to 10 rad/sec). To make the visual and motion cues correspond to each other, a 0.1-second delay was placed in the visual loop. All the simulations of vehicle dynamics included this delay. As shown in figures 1 and 2, the subjects were in a compensatory system. They had the task of processing visual and roll-motion cues of system error, along with their knowledge of the vehicle dynamics, to manipulate the control stick so as to stabilize and control the man-vehicle system.

Paralleling the work of McRuer et al. (ref. 2), data were taken during 89.6-second runs: the human operator's input and output were sampled every 0.1 second. The data taken were of display and control-stick voltages which were scaled so as to be directly comparable. The data were processed in the same manner as by McRuer et al., except that a hybrid computer was used instead of watt-hour meters. The actual calculations performed (excepting further calculations to correct for the remnant) were

¹The system input was the sum of 10 sinusoids. The higher frequency sines were reduced in amplitude to one-tenth the amplitude of the lower frequency sines. The frequency at which the reduction in amplitude occurs is called the input breakpoint. See references 2 and 3.

$$|Y_p(\omega_k)| \doteq \frac{|M_m(\omega_k)|}{|e(\omega_k)|} \quad \text{amplitude ratio of the human operators} \quad (4)$$

describing function

$$\angle Y_p(\omega_k) = \angle M_m(\omega_k) - \angle e(\omega_k) \quad \text{phase of the human operators} \quad (5)$$

describing function

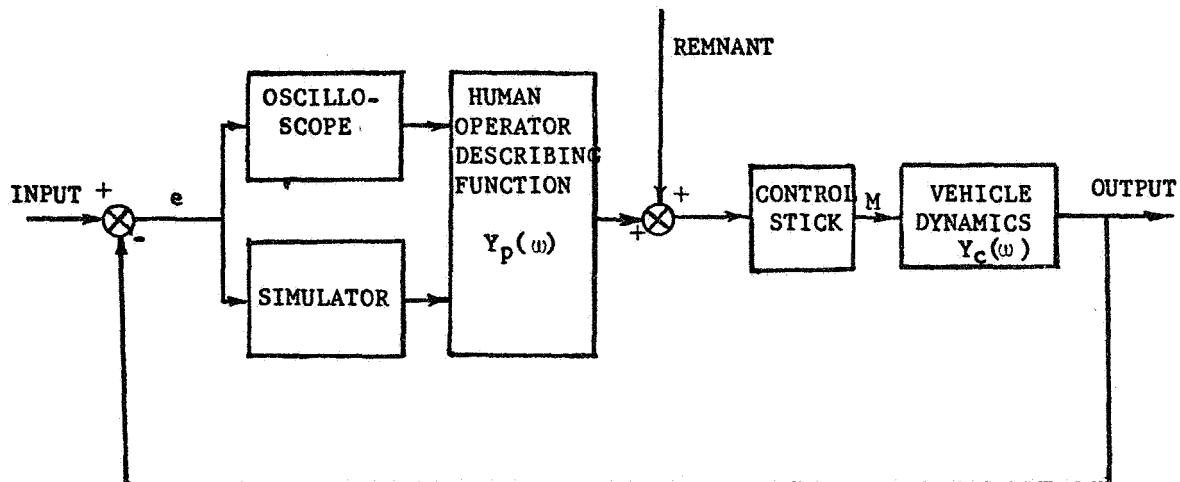


Figure 1. —The human operator in a compensatory system.

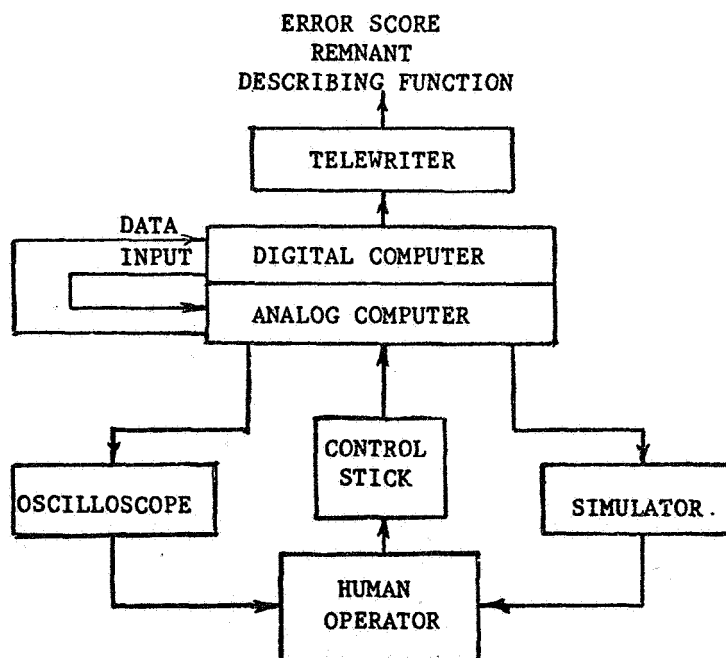


Figure 2. —The experimental set-up.

where

$$\begin{aligned} |M_m(\omega_k)|^2 &= (A_{mk}^2 + B_{mk}^2)(\Delta t)^2 & \angle M_m(\omega_k) &= \tan^{-1} \left(\frac{B_{mk}}{A_{mk}} \right) \\ |e(\omega_k)|^2 &= (A_{ek}^2 + B_{ek}^2)(\Delta t)^2 & \angle e(\omega_k) &= \tan^{-1} \left(\frac{B_{ek}}{A_{ek}} \right) \end{aligned}$$

where

$$\begin{aligned} A_{mk} &= \sum_{n=1}^{896} M(n\Delta t) \sin(\omega_k n\Delta t) & A_{ek} &= \sum_{n=1}^{896} e(n\Delta t) \sin(\omega_k n\Delta t) \\ B_{mk} &= \sum_{n=1}^{896} M(n\Delta t) \cos(\omega_k n\Delta t) & B_{ek} &= \sum_{n=1}^{896} e(n\Delta t) \cos(\omega_k n\Delta t) \end{aligned}$$

where $e(n\Delta t)$ and $M(n\Delta t)$ are the sampled values of the man's input and output as shown in figure 1. The system input was the sum of 10 sinusoids of known amplitude, frequency (the ω_k 's), and phase in a "shelf" spectrum. See reference 3 for further details of the experiment as well as a complete presentation of the experimental data.

Figure 3 shows typical experimental results for one set of vehicle dynamics, namely for

$$Y_c(s) = \frac{e^{-0.1s}}{s}$$

The data include the human operator's describing function (amplitude ratio and phase versus frequency), the human operator's remnant (power spectral density of that portion of the human operator's output which is uncorrelated with his input), and system relative integral squared error (the ratio of the average input power during the runs), all for conditions of visual cues only, roll-motion cues only, and simultaneous visual and roll-motion cues.

THE REMNANT CORRECTION

There has been some controversy as to how the measurements of the human operator's describing function are affected by the remnant. Some claim that in calculating the human operator's describing function, using the cross-correlation functions between the system input and the human operator's input and between the system input and the human operator's output, the effects of the remnant are eliminated. Unfortunately this is not completely true: the cross-correlation technique tends to reduce the effect of the remnant on the measurements,

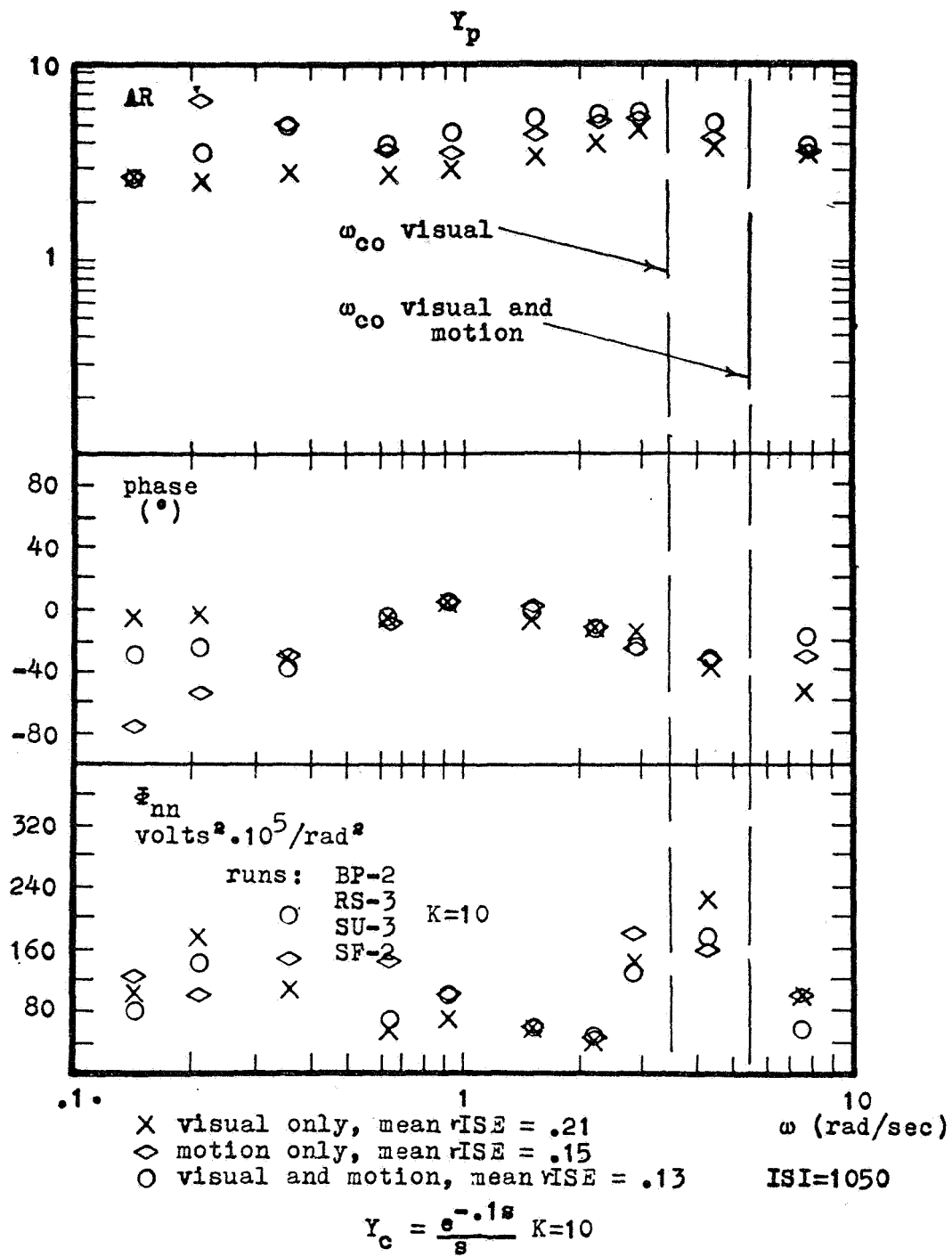


Figure 3.—Experimental results.

but does not eliminate it. Because data are taken for a finite run time, all the measurements have a finite bandwidth, and some remnant enters into all the calculations. The pertinent question, then, is not whether the remnant affects the measurements of the human operator's describing function, but rather how much the remnant affects the measurements. In reference 3 an analytical expression for the effects of the remnant on the experimental measurements of the human operator's describing function was derived. The expression permitted correction of the data for the effects of the remnant as shown in figure 4. As can be seen, the effects of the remnant are small compared to one standard deviation of the data and hence may be ignored. This held true for all the data obtained.

RESULTS

The data were fitted with models of the type

$$Y_p(s) = \frac{Ke^{-\tau d s}(\tau_1 s + 1)(\tau_2 s + 1)}{(\tau_3 s + 1)(\tau_4 s + 1)(\tau_5 s + 1)} \quad (6)$$

or of the type

$$Y_p(s) = \frac{Ke^{-\tau d s}(\tau_1 s + 1)(s^2 + 2\xi\omega_n s + \omega_n^2)}{\omega_n^2(\tau_2 s + 1)(\tau_3 s + 1)} \quad (7)$$

A brief discussion of these models is given in reference 3. The model given by equation (6) is developed in reference 2. Examination of the values of the model parameters at each point for which data were taken, as well as direct examination of the data, leads to the following conclusions about the human operator's use of roll-motion cues in man-vehicle control. As shown schematically in figure 5, when roll-motion cues are added to visual cues, the human operator increases his phase lead at the higher frequencies (generally above 2 rad/sec). This in turn permits him to increase his gain, and hence his crossover frequency,² without a loss in phase margin. These changes in the human operator's describing function occur very consistently throughout the data as shown in table 1.

Since roll-motion cues are of help to the human operator primarily at the higher frequencies, one would expect motion cues to be most beneficial for vehicle dynamics which lead to more high-frequency motions (motions above 2 rad/sec). Figures 6 through 8 show that such is the case. In figures 6 and 7 we see that low-order vehicle dynamics, which pass more high-frequency signals, permit the human operator to make greater use of roll-motion cues when these roll-motion cues are added to visual cues. "Greater use" is measured as the percent improvement in the integral squared error during a run, when the roll-motion cues are added to the visual cues. Furthermore, figure 6 shows that for vehicle dynamics with a high control-stick gain, which permits the human operator to force the vehicle to make rapid responses, the human operator can make more use of roll-motion cues than for the same vehicle with a lower control-stick gain. Finally, figure 8 shows that without exception (in the range of vehicle dynamics studied), the addition of a pole at the origin to a second order system leads to the human operator being less able to utilize roll-motion cues.

²The crossover frequency is that frequency at which the open loop gain (the product of the magnitudes of the human operator's describing function and the vehicle dynamics) passes from greater than unity to less than unity.

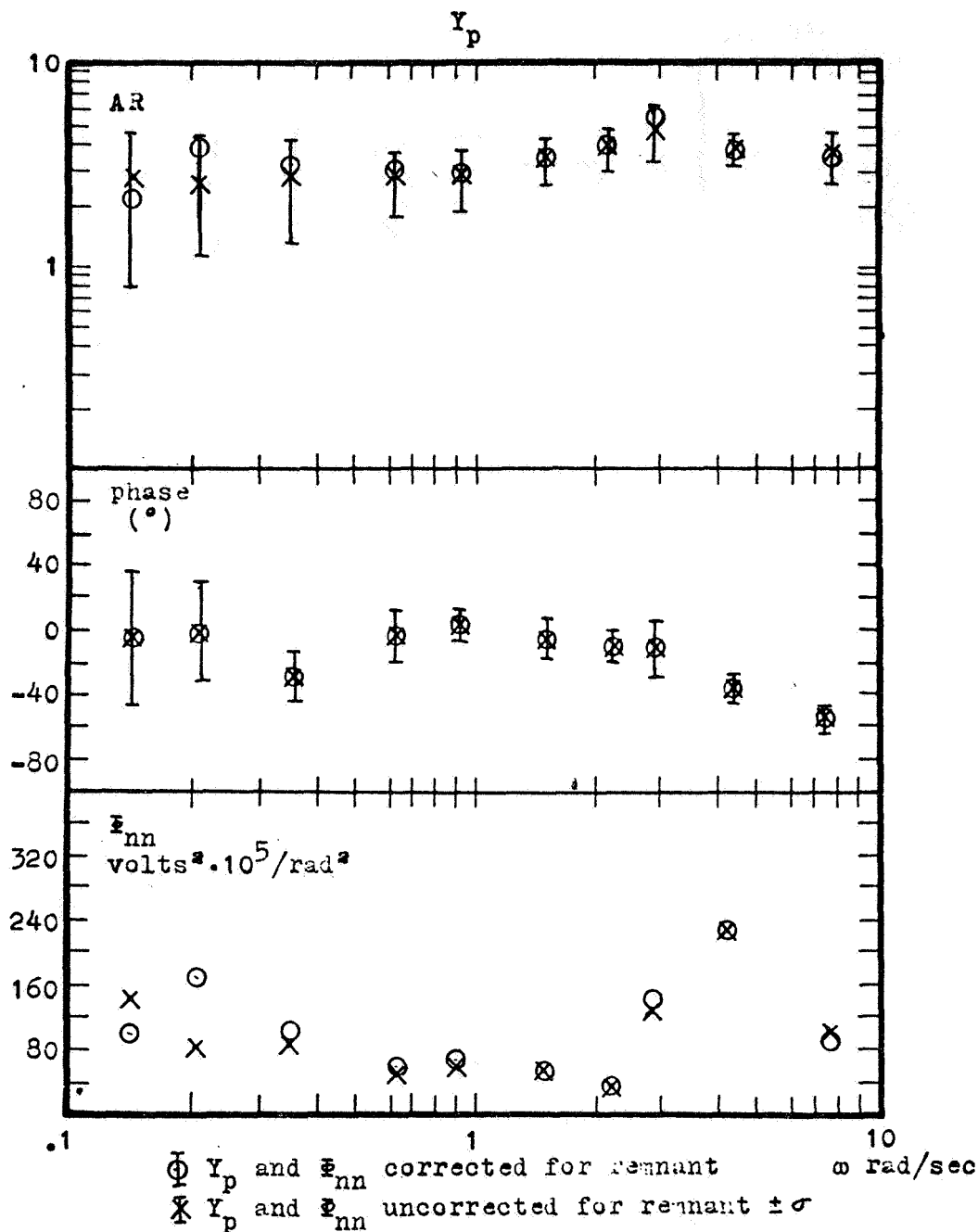


Figure 4. —The effect of the remnant correction to Y_p for $Y_c = e^{-.1s}/s$.

It should be remembered that these results are not based on a comparison of error scores, but on a comparison of percentage improvements in error scores when roll-motion cues are added to visual cues. It is interesting to note that at no point did the addition of roll-motion cues actually degrade performance, although there were dynamics for which motion cues were of no help to the human operator, such as for

$$Y_c(s) = \frac{2e^{-.1s}}{s^2(s+2)}$$

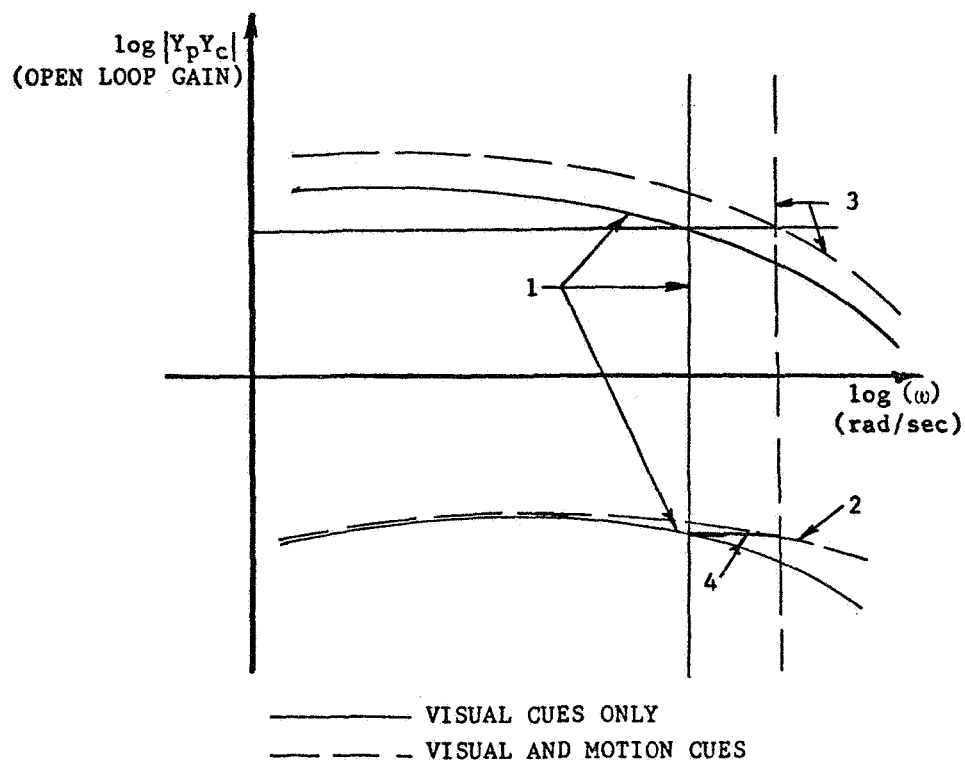


Figure 5. — How the human operator changes his control behavior upon the introduction of roll motion cues. (1) While controlling a vehicle with visual cues only, the human operator-plant combination has a certain amplitude ratio, phase, and cross-over frequency. (2) The addition of roll motion cues to visual cues of system error permits the human to generate more lead at the higher frequencies. (3) Thus he can increase the system gain (and the cross-over frequency). (4) Without a loss in phase margin.

It should also be noted that figure 8 confirms the finding of reference 1 that motion cues help most for unstable vehicle dynamics which are near the limits of controllability.

SUMMARY

The objective of this investigation was to learn about the human operator's use of roll-motion cues and, thus, to take a step in the direction of predicting actual flight data for the human operator from fixed-base laboratory experiments. In reference 4, Meiry has shown that the vestibular dynamics are very similar in pitch and roll. Thus it would seem appropriate to assume that the results obtained with roll-motion cues are similar to, if not the same as, the results which would be obtained if pitch-motion cues were used. Therefore, we can say that when the differences between fixed-base laboratory data and flight data for the human operator are due to the absence of pitch- and roll-motion cues in the

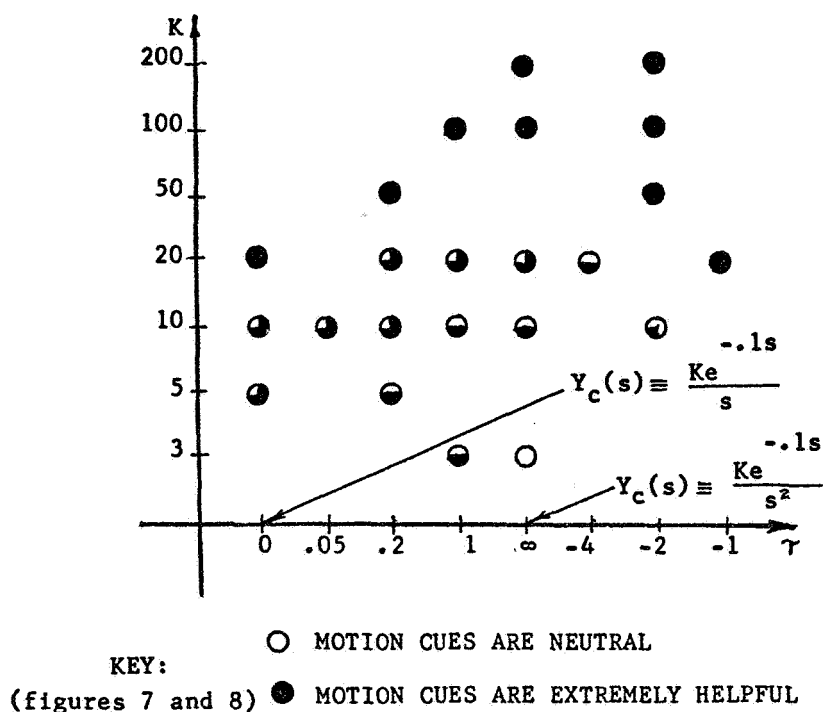


Figure 6.—The effects of motion cues on the performance of the man-vehicle system for

$$\text{the vehicle dynamics } Y_c(s) = \frac{Ke^{-.1s}}{s(\tau s + 1)}.$$

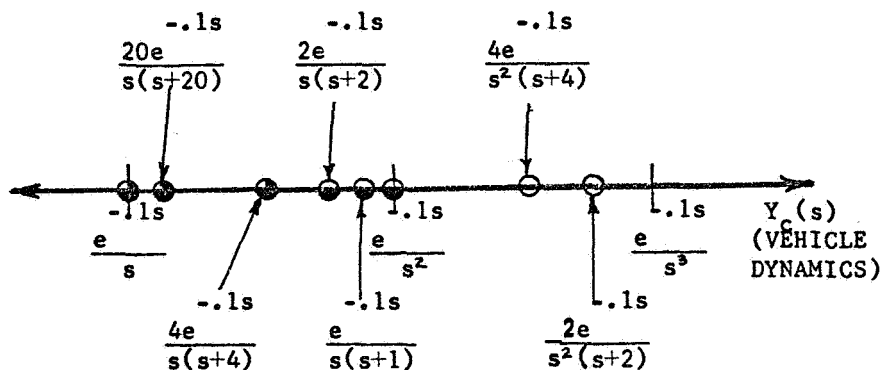


Figure 7.—The effects of motion cues on the performance of the man-vehicle system as the vehicle dynamics range from $e^{-.1s}/s$ to $e^{-.1s}/s^3$.

laboratory, the data given in reference 3 permit prediction of flight data from fixed-base laboratory data for the human operator. These predictions are quantitative, and not merely qualitative in nature. They are, however, restricted to vehicle dynamics for which data were taken, namely equations (1) through (3). It is possible to estimate results for vehicle dynamics of the types

$$Y_c(s) = \frac{(1+Cs)}{s^2} \quad (8)$$

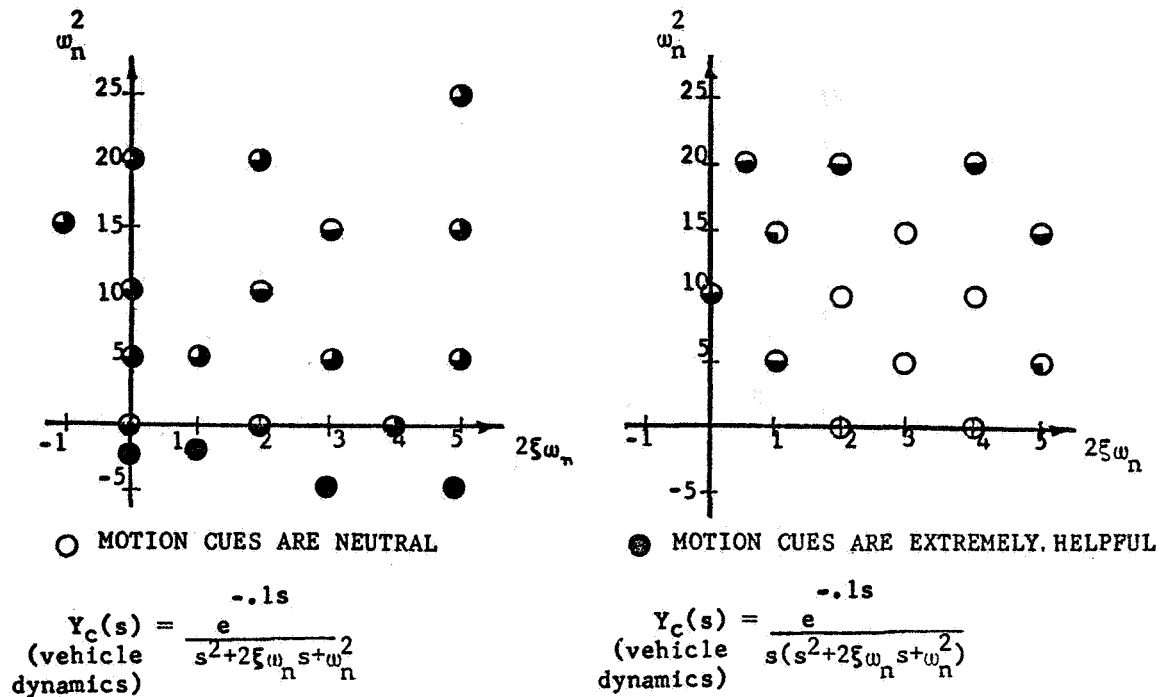


Figure 8. —The effects of motion cues on the man-vehicle system for various vehicle dynamics.

and

$$Y_c(s) = \frac{(1+Cs)}{s(s^2 + 2\xi\omega_n s + \omega_n^2)} \quad (9)$$

by interpolating between results for which data were taken.³

While the results presented here should be very useful, much remains to be done. At present, experiments are underway at MIT to learn how the human operator uses yaw-motion cues, and at NASA-ERC to measure the human operator's describing function over a wider range of dynamics. Experiments are planned to study the human operator's use of translational motion cues. Finally, a physiologically based model of the human's use of motion cues, using our knowledge of vestibular dynamics, remains to be developed.

³For example, data were taken for the vehicle dynamics $Y_c(s) = e^{-.1s}/s$ and $Y_c(s) = e^{-.1s}/s^2$ (eq. (2)). As C varies from ∞ to 0 equation 8 varies from $Ke^{-.1s}/s$ to $Ke^{-.1s}/s^2$, between the two points for which data were taken.

REFERENCES

1. Young, L. R.: Some Effects of Motion Cues on Manual Tracking. J. Spacecraft Rockets, Oct. 1967, pp. 1300-1303.
2. McRuer, D.; Graham, D.; Krendel, E.; and Reisener, W., Jr.: Human Pilot Dynamics in Compensatory Systems. Tech. Rept., no. AFFDL-TR-65-15, Wright-Patterson Air Force Base, July 1965.
3. Shirley, R. S.: Motion Cues in Man-Vehicle Control. Man-Vehicle Control Lab. Rept., no. MVT-68-1, Feb. 1968 (ScD thesis, MIT).
4. Meiry, J. L.: The Vestibular System and Human Dynamic Space Orientation. MIT Man-Vehicle Control Lab. Rept., no. T-65-1, June 1965. (ScD thesis, MIT).

VII. DEVELOPMENTS IN CONTROL AND DISPLAY SYSTEM DESIGN

27. Manual Control System Performance With Predictive Displays*

*John DeShon Warner
Boeing Company*

An experimental study is reported in which predictive displays were evaluated as a performance aid in a representative manual control task. Predictive information reduced the task complexity to the point where operator performance remained consistent at a high level, though expected variations in task-oriented performance criteria were found as a function of several parameters of the controlled element.

INTRODUCTION

The predictive display technique has been receiving considerable attention since 1960 as an effective aid in certain manual control problems. Based on a research program, of which the experimental effort is reported here, it appears that applications which are most likely to benefit from use of the predictive technique fall into one or more of the following categories:

- (1) The dynamics of the controlled element are complex and slowly responding.
- (2) The dimensions of the control task or number of individual tasks are relatively large.
- (3) The nature of the task requires considerable anticipation by the operator.
- (4) Optimization of some cost function is the primary objective of the mission.
- (5) The task is time-constrained or is a terminal control task.
- (6) There is a strong desire for flexibility in the mission.

Previous studies of the technique have indicated several potential advantages to its use in manual control systems:

- (1) Learning times can be decreased.
- (2) Human-operator effectiveness in terminal control tasks can be improved.
- (3) Manual control can approach optimal control with respect to a specified performance criterion.
- (4) Control of nonlinear systems and of linear systems with pure time delays and other nonminimum phase characteristics can be improved.
- (5) The operator can plan optional courses of action to increase the likelihood of mission success.

*This research was sponsored in part by the National Aeronautics and Space Administration under contract NASr 54(06) while the author was at the University of Michigan.

(6) Information processing requirements on the human operator can be reduced, especially in multidimensional control tasks.

Despite all this, predictive displays have not yet found their way into operational vehicles. In some instances this may be due to hardware considerations, but it is more likely a result of a lack of basic knowledge of the technique and its advantages and limitations. In any application, predictive instruments will have to be weighed against other advanced control and display concepts. Nevertheless, because of their potential, there is considerable motivation for conducting general studies aimed at increasing our understanding of predictive displays.

To provide a starting point for both general studies and applications of predictive displays, a research program was instituted with the purpose of clarifying the applicable terminology and providing a comprehensive discussion of the inherent characteristics of the technique. This paper presents the experimental portion of that study, which was designed to answer the following questions:

(1) For a representative manual control system, under what circumstances will the predictive display provide a performance improvement over nonpredictive displays?

(2) Using a predictive display, how can performance variations be explained for changes in the parameters of the controlled element?

EXPERIMENT

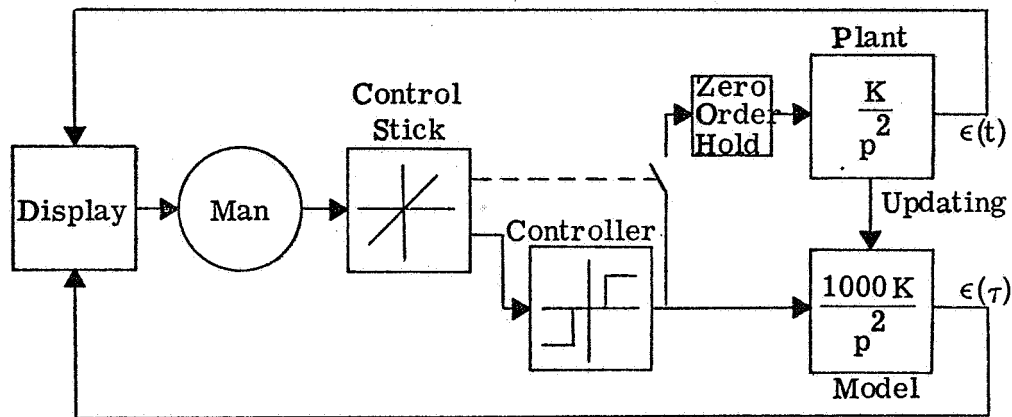
To eliminate certain variables from the study, the fast-time model method of Ziebolz and Paynter for generating predictive information was chosen. This technique has the advantages of relative simplicity in mechanization and (provided the controlled element and controller dynamics and future disturbances are completely known) the capability for supplying totally accurate predictions.

A minimum-time terminal control task for a pure inertia system driven by a fixed three-state relay controller was used in the experiment. The desired terminal state was specified to be zero error and error rate. The controlled element gain K and initial conditions on the error and error rate were varied to achieve a systematic variation in several independent parameters discussed below.

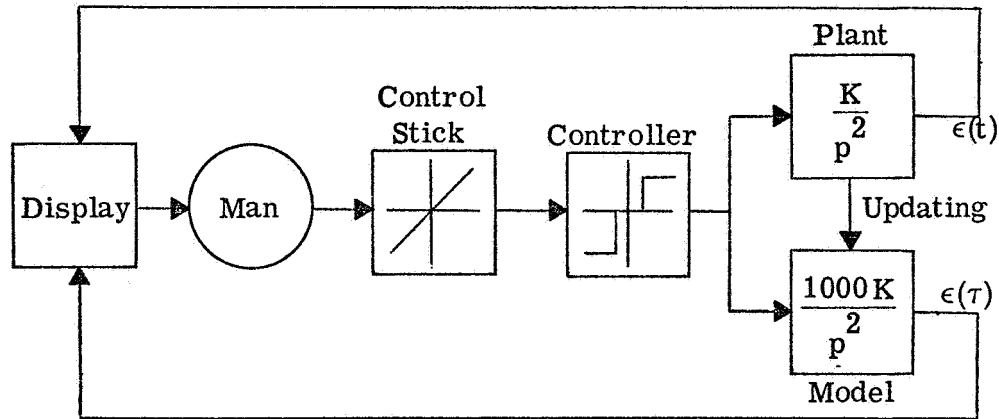
Three display forms were chosen for evaluation: exploratory prediction, online prediction, and normal (no prediction) in a time-versus-error format. The fast-time model implementation schemes used for the predictive displays are illustrated in figure 1, and the display format in figure 2.

The subjects were instructed to drive the system output to the specified terminal state and then maintain that state for a finite time interval. Several new performance measures were developed for this task. Discussion here will be restricted to the minimum time-to-go (MTTG) criterion, which measures the minimum time from the state that actually exists at a prespecified terminal time (the true minimum time based on the gain and initial conditions) to the desired final state. This is a task-oriented criterion which indicates the level of success of the transient portion of the task. To provide an index of human-operator performance, the timing errors (Δt_1) in application of the first control reversal were measured. Recall that the time-optimal control of a pure inertia system follows the well-known bang-bang control law in which maximum effort is applied towards the target followed by a reversal in control at t_1 for maximum deceleration so that the target is reached just as the speed becomes zero. A lag in control reversal yields a positive value of the measure Δt_1 . These two criteria are illustrated in figure 3.

Preexperiment analysis, including a pilot study, indicated several independent parameters which might be influential to performance. The most significant of these



(a)



(b)

Figure 1.—Predictive display mechanization. Model time scale = 1000 times real time.
(a) Exploratory prediction, (b) online prediction.

parameters is termed the switch time t_{sw} . It is dependent upon both the gain and the initial conditions and is equal to the time remaining to reach the target after the decelerating input is applied, assuming optimum control reversal. Switch time is illustrated in figure 3. With a predictive display, the prediction span must be at least equal to the switch time in order that all required mental predictions of the response be eliminated. Similarly, with the normal display indicating only the present state, the switch time is the required time span over which the operator must predict the system response for the decelerating input.

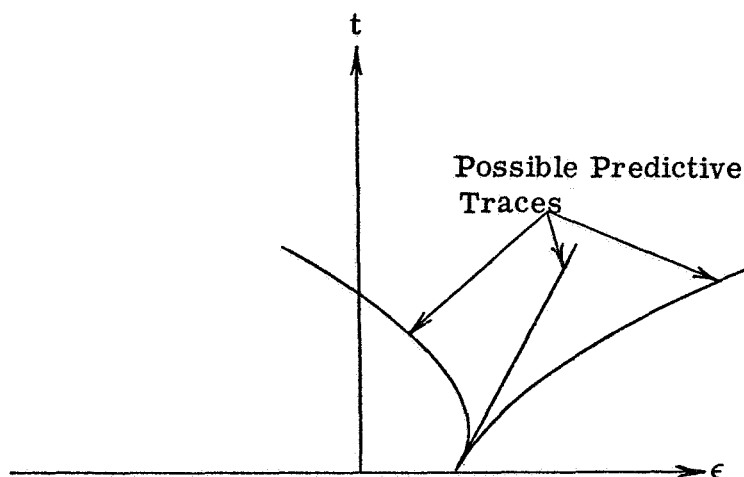


Figure 2. —Display format at the beginning of a trial.
(Prediction span = 7 secs, repetition rate = 30 cps.)

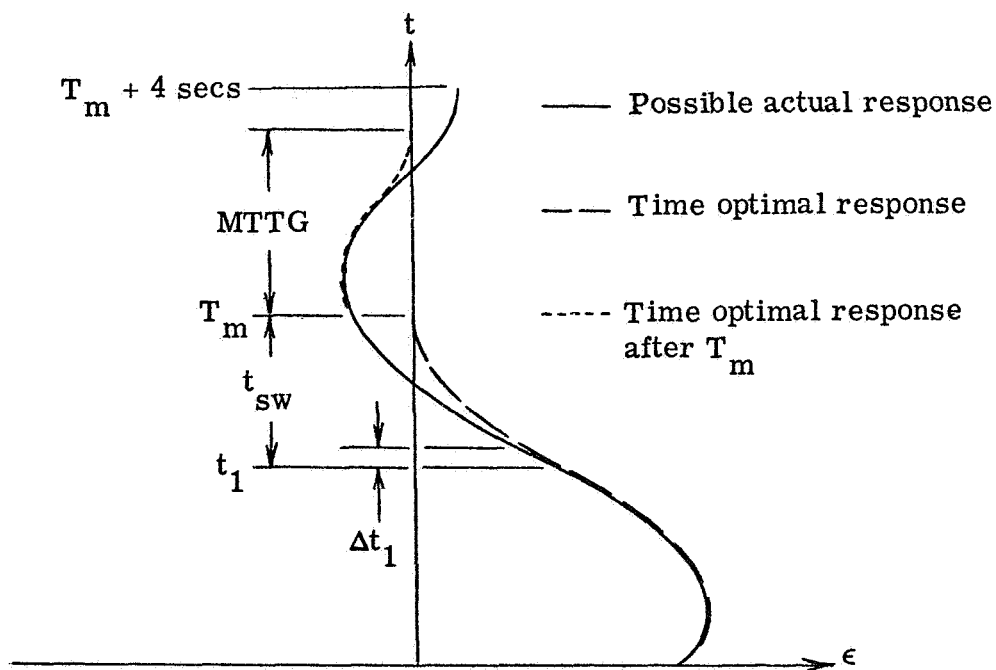


Figure 3. —Performance measures and the independent parameter, switch time.
(T_m = true minimum time, t_1 = optimum time of control reversal.)

RESULTS

The experiment was conducted using five volunteer subjects who had no prior experience in this type of a task. The subjects were seated in front of a large-screen oscilloscope with their right hand on a control stick which could be displaced laterally. A trigger on the control stick was used for operation with the exploratory predictive display.

Each subject performed for 13 1-hour sessions. All display forms and combinations of independent variables were presented each session to avoid confounding of the results with learning effects. Each session was divided into three blocks for three different gains, with twelve separate conditions for each gain. Display ordering within each block was always exploratory prediction first, followed by online prediction, and, finally, the normal display. This allowed the maximum benefit from short-term learning to accrue to the nonpredictive display.

Statistical testing showed that the effects of learning were insignificant over the last six sessions of the experiment. Results for each combination of display and condition were then averaged across these six sessions and the five subjects.

The measure of the timing errors in application of the first control reversal yielded an interesting result for exploratory prediction: both the average values of Δt_1 and its standard deviation were insensitive to any of the independent parameters despite a concerted effort to design conditions which would make operation with that display difficult. On the other hand, the timing errors for the normal display showed a large variability which increased with increases in t_{sw} and a marked tendency toward early switching. (It can be shown that the performance loss for a late control reversal is significantly greater than that for an equally early control reversal.) Representative distributions of the Δt_1 measure for the two displays are shown in figure 4.

A performance variation for the predictive displays was found in terms of t_{sw} . This can be explained by analyzing the sensitivity of the minimum time-to-go criterion to a constant timing lag Δt_1 . This sensitivity is given by

$$MTTG = 2\Delta t_1 \left[1 + \sqrt{1 + \frac{2t_{sw}}{\Delta t_1}} \right]$$

Figure 5 shows the averaged minimum time-to-go data against t_{sw} for exploratory prediction, with a curve superimposed that is described by the above equation with a Δt_1 lag of 0.02 second. The results for online prediction are shown in figure 6. Only a few of the conditions yielded statistically significant differences between online and exploratory prediction.

A more significant variation with t_{sw} was shown for the normal display, as illustrated in figure 7. Statistical tests indicated that the differences between the predictive displays and the normal displays were significant only at the higher values of switch time.

CONCLUSIONS

The predictive displays were seen to have nearly eliminated variability of operator performance. This suggests that performance without the predictive information is almost entirely a function of the predictive capabilities of the operator. This conclusion is further demonstrated by the deterioration of performance with increases in the required mental prediction time span (t_{sw}) when operating with the normal display.

It was shown that performance variations with the predictive displays are dependent upon the nature of the controlled element and the performance measure itself. While such variations may then seem to be artifacts of the independent parameters and of the criteria, they are nevertheless very real.

The exploratory predictive display gave only marginal improvements over online prediction. However, it should be noted that this conclusion should be confined to situations

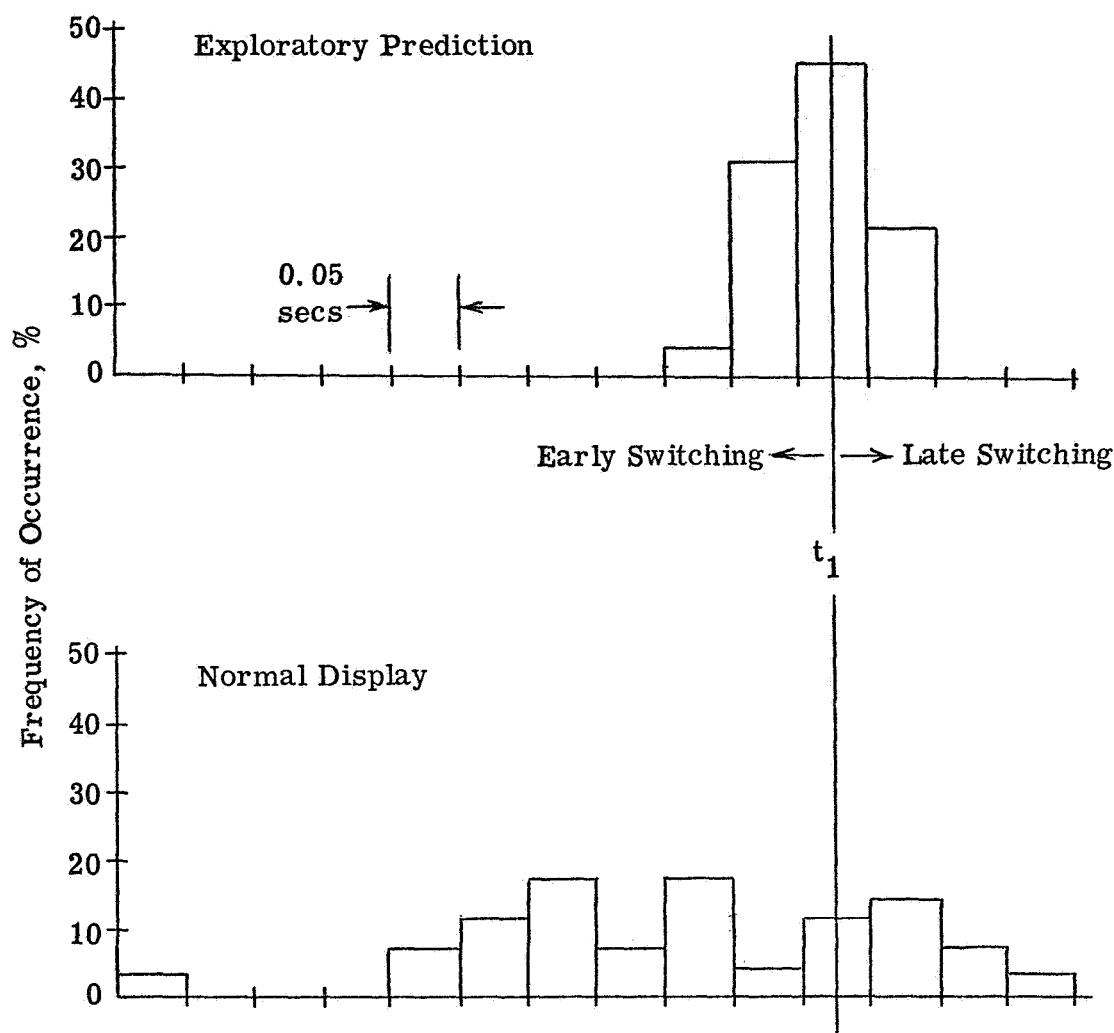


Figure 4.—Typical distributions of timing errors in application of first control reversal.

in which the required decision times are relatively short. If some appreciable searching for the proper control inputs was required, it can be expected that the online predictive display would cause significant performance penalties beyond that obtainable with exploratory prediction.

Based on experience gained in this study and conversation with the subjects during testing, a serious question arose concerning a predictive display as a training aid. Predictive information provided early indications to the subjects that a bang-bang control was required for optimum performance. Once that was learned however, it is somewhat doubtful that the predictive displays were of further assistance to operation with the normal display. In the latter situation the operator must process the current error and inferred error-rate information and form a prediction based on this information. With prediction however, his attention is focused only on the predictive path in the vicinity of the target. Since the predictive cues are thus altogether different from the nonpredictive cues, it is not apparent how prediction could have materially altered the learning curve for the normal display. Further studies in this area are definitely needed.

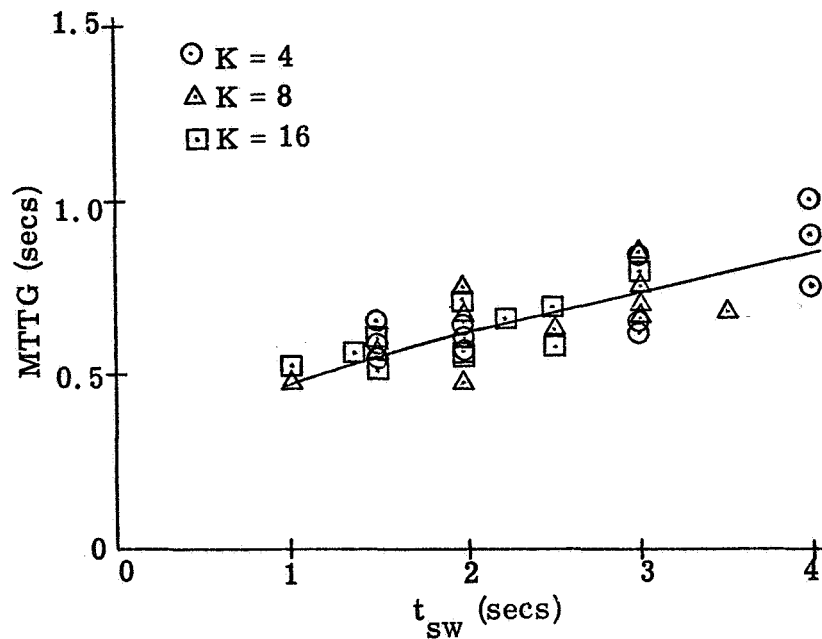


Figure 5.—Average MTTG performance measure vs. switch time for exploratory prediction.

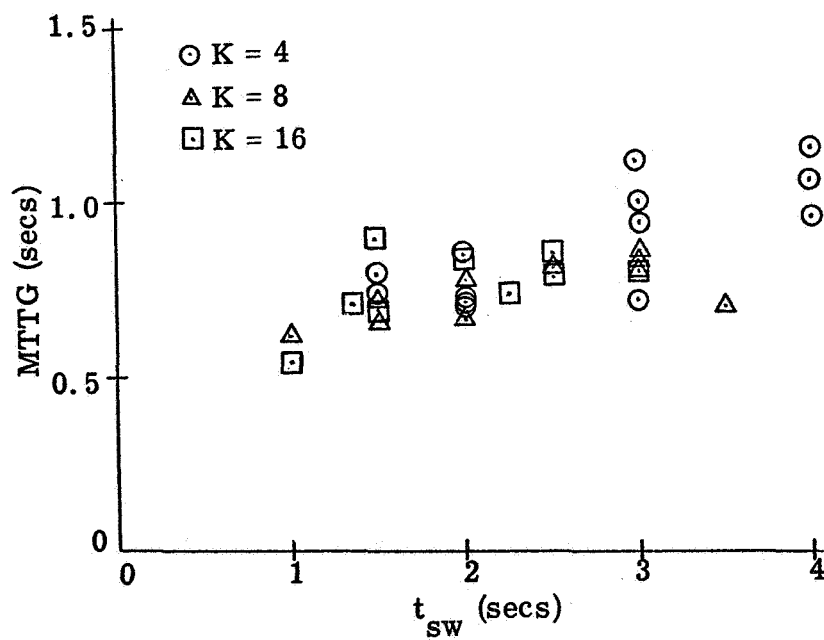


Figure 6.—Average MTTG performance measure vs. switch time for online prediction.

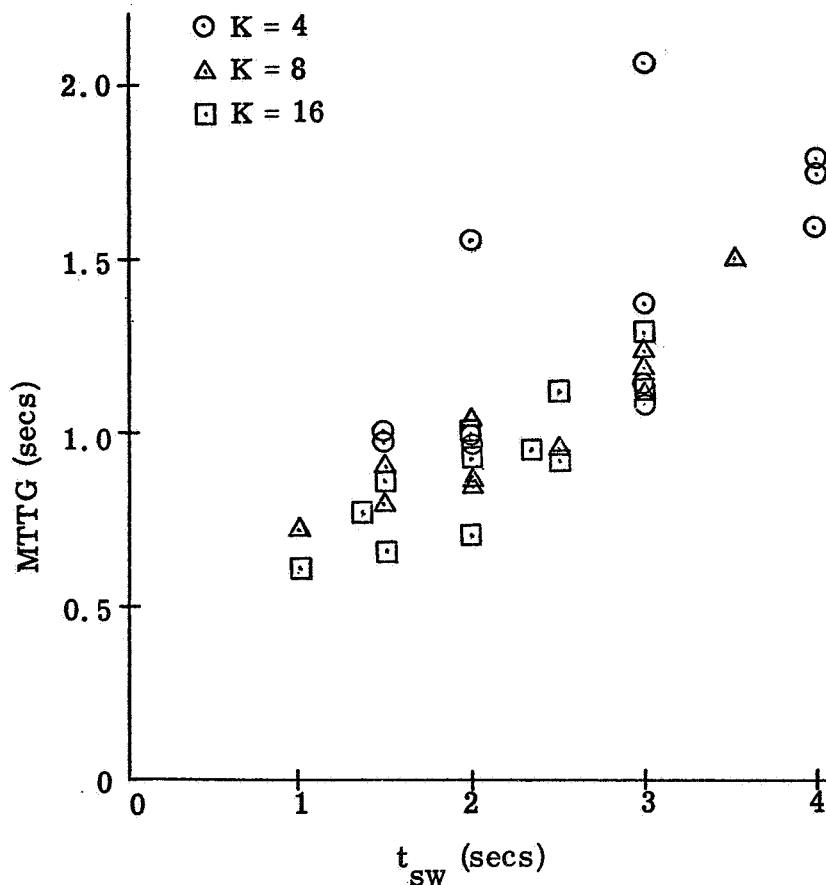


Figure 7. —Average MTTG performance measure vs. switch time for the normal display.

This experiment was designed with the hope that the results could be extended to other types of systems. However, any extensions should be made only with the following constraints in mind:

(1) The number of alternative control actions available to the operator must be such that the decision process can be considered to be discrete rather than continuous.

(2) The time available to make a control decision and action is greater than the time required by the operator.

(3) The predictive display format and predictor control program (which supplies the input to the fast-time model) are such that no additional mental predictions are required by the operator to effect a nearly optimal control law.

RECOMMENDATIONS

Predictive displays hold perhaps the most promise in situations where the dynamics of the vehicle and required control variations are quite complex. In such circumstances it is either very difficult or even impossible to provide accurate predictions over a long timespan. It would be of considerable practical interest to investigate the utility of predictive information supplied by fast-time models which represent gross simplifications of the actual system being controlled.

Still needed are quantitative, general studies of the effects of prediction span and repetition rate. In addition, basic research is needed on display formats and implementation schemes of complex predictor control programs that provide hypothetical inputs to the fast-time model.

Finally, to more fully understand the value of predictive information, further research is needed on the question of how the human operator predicts the response of a system.

REFERENCES

1. Bernotat, R.: Das Prinzip der Voranzeige und Seine Anwendung in der Flugführung. Zeitschrift für Flugwissenschaften, vol. 13, Oct. 1965, pp. 373-377.
2. Kelley, C. R.; Mitchell, M. B.; and Strudwick, P. H.: Applications of Predictor Displays to the Control of Space Vehicles. Dunlap & Associates, Inc., Technical Report to the NASA Office of Manned Space Flight, Apr. 1964.
3. Kelley, C. R.: Further Research on the Predictor Instrument. Dunlap & Associates, Inc., Technical Report 252-60-2 to the Office of Naval Research, Dec. 1960.
4. McCoy, W. K., Jr.; and Frost, G. G.: Predictor Display Techniques for On-Board Trajectory Optimization of Rendezvous Maneuvers. AMRL-TR-66-60, May 1966.
5. Sheridan, T. B.: Three Models of Preview Control. IEEE Transactions on Human Factors in Electronics, vol. HFE-7, no. 2, June 1966, pp. 91-102.
6. Warner, J. D.: A Fundamental Study of Predictive Display Systems. Ph.D. thesis, Univ. Mich., Dec. 1967.
7. Wingrove, R. C.; and Coate, R. E.: Piloted Simulation Tests of a Guidance System Which Can Continuously Predict Landing Point of a Low L/D Vehicle During Atmosphere Re-Entry. NASA TN D-787, Mar. 1961.
8. Ziebolz, H.; and Paynter, H. M.: Possibilities of a Two-Time Scale Computing System for Control and Simulation of Dynamic Systems. Proc. Natl. Electronic Conf. vol. 9, Feb. 1954, pp. 215-223.

28. Some Relations Between Visual and Kinesthetic Displays in Normal Driving*

*Malcolm L. Ritchie
Ritchie, Inc.*

Where drivers have a free choice of speeds in curves they choose the speeds in such a way that lateral acceleration is an inverse function of forward velocity. An expression is developed which fits the data in allowing for individual differences in g-levels and a common slope.

INTRODUCTION

Every student of manual control behavior is familiar with the class of display variables called "feel of the road" in driving and "seat of the pants" in flying. The hypotheses which are explored in this paper are that a significant portion of this little-understood class of variables is the direct sensing (kinesthetic perception) of acceleration, and that this kinesthetic display is an important determiner of manual vehicle control behavior.

Let us consider a driver making a cross-country trip on roads such that he has to slow down to take the curves. It is normal for the driver to slow down to his desired speed before entering the curve. He looks the curve over and decides that a particular speed is appropriate. The choice is, of course, based upon previous experience, but the focus at the moment is on the criterion, or the criteria, for that choice. One possibility is that he tends to adjust speed below the speed limit so as to take all curves at the same turn rate, or

$$\text{Turn rate} = \text{constant} \quad (1)$$

If the turn rate is constant, then the following relation would also be true, that

$$g = KV, \quad (1A)$$

where g is lateral acceleration, K is a constant, and V is the forward velocity.

Another possibility is that the driver will adjust his speed so as to produce the same lateral acceleration.

$$g = K \quad (2)$$

where g is lateral acceleration and K is a constant.

*The research upon which this publication is based was performed pursuant to contract No. PH-86-67-155 with the Public Health Service, Dept. of Health, Education and Welfare.

Still another possibility is that the lateral acceleration is a major criterion for his choice of speed but the criterion g decreases as speed increases. This relationship may have the following form:

$$g = C - KV \quad (3)$$

where g is lateral acceleration, C is a constant, K is another constant, and V is the forward velocity. C may represent the personal confidence in the system to tolerate g , and K an estimate of increasing danger.

These hypothetical relations are simple but distinct, and we should be able to decide if a set of data fit one or the other, at least as a class.

THE EXPERIMENT AND RESULTS

The data upon which the present discussion is based are reported in part in a paper which will appear in *Human Factors* (ref. 1). The experiment involved driving over a section of Ohio highways containing many curves. The critical variable is the choice of speeds in the curves. The experimental conditions were designed to give the subject a free choice of the speed at which he took each curve. The instructions were to drive as he normally did on a cross-country trip. In each curve the experimenter actuated a recording of (1) the car's speed and (2) the lateral acceleration as measured by an accelerometer mounted on the floor of the car under the driver's seat. Twenty-five male and twenty-six female subjects drove the 110-mile course, producing 10 290 observations.

When lateral acceleration was plotted against forward velocity, a strong inverse relation was obtained for all speeds except those below 20 mph. Figure 1 shows this relationship for men versus women. It appeared that the difference between men and women might be accounted for by the fact that the men drove faster on the average than did the women. To explore this possibility the subjects were ranked by overall speed and divided into 5 groups with the 10 fastest in group 1, the next 10 fastest in group 2, etc. The data plotted in this way appear in figure 2.

Now, if we refer back to the equations at the beginning of our discussion, we will see that the results generally follow equation (3). By making some limiting assumptions we can come up with a more refined relationship. First, let me leave off the values below 20 mph. We can assume that at very low speeds drivers are quite limited in the accelerations they produce. We can get support from this position from the data, which show no differences among groups below 20 mph. Next, we may leave off the values above 60 mph. The speed limit over most of the course was 60, and a different set of variables may be introduced thereby.

Using these limitations, an expression has been developed to fit the observed data. It is

$$g = P - 0.168 \left(\frac{v}{20} - 1 \right) \quad (4)$$

where g = lateral acceleration in gravitational units, P = a personal variable constant expressing g -tolerance, and v = forward velocity in miles per hour between 20 and 60. The fit is approximate. About 90 percent of the P -values lie between 0.33 g and 0.19 g .

One of the interesting things about equation (4) is the variable constant P . A part of its constituents seems to be the confidence the individual has in the car-driver-roadway system. For this reason it can be expected to remain constant over a course such as the

one in this experiment. Another component of P is the level of aspiration. This component may also be expected to vary little through the course on a given day, but may change markedly from day to day or from one driving condition to another.

One possible source of error in interpreting the findings of the reported experiment is that the curve signs, advisory speed signs, or other features of the normal highway complex may have determined the speed for the driver. As a partial test of this hypothesis a study was made of the 57 curves on the course which contained advisory speed signs. The results of this analysis are shown in figure 3. This figure shows that the drivers drove faster than the advisory sign speeds below 40 mph, but above 40 did not exceed the posted speed. This suggests that the lateral- g -speed relationship of equation (4) held in spite of the advisory speeds. A preliminary calculation suggests that the constant 0.068 was reduced by the presence of the signs by perhaps 35 percent. Future reports will expand upon this issue.

The definite trend to the speed-acceleration function suggests that a lateral acceleration may indeed be a primary criterion for the choice of speed in curves. It is possible that this value is sensed visually as the degree of lean of the auto body relative to the roadway surface. It seems more likely that the display of the lateral acceleration force is kinesthetic.

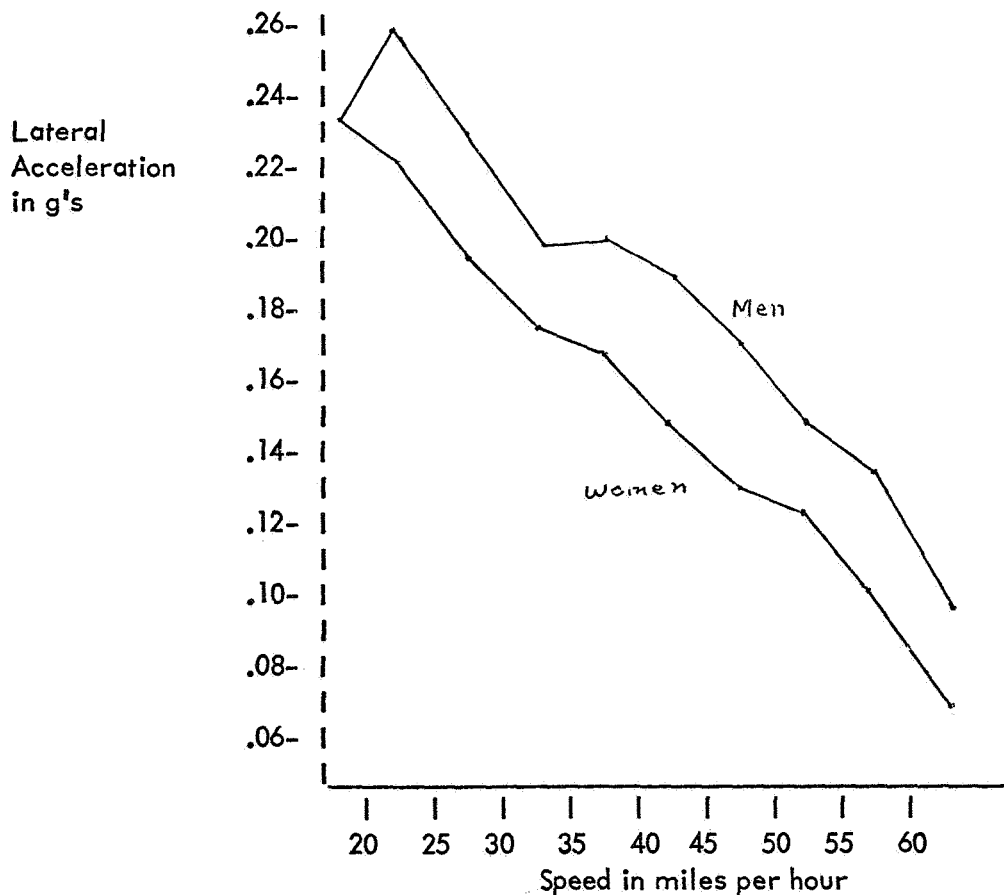


Figure 1. —Lateral acceleration as a function of speed in curves for men versus women subjects.

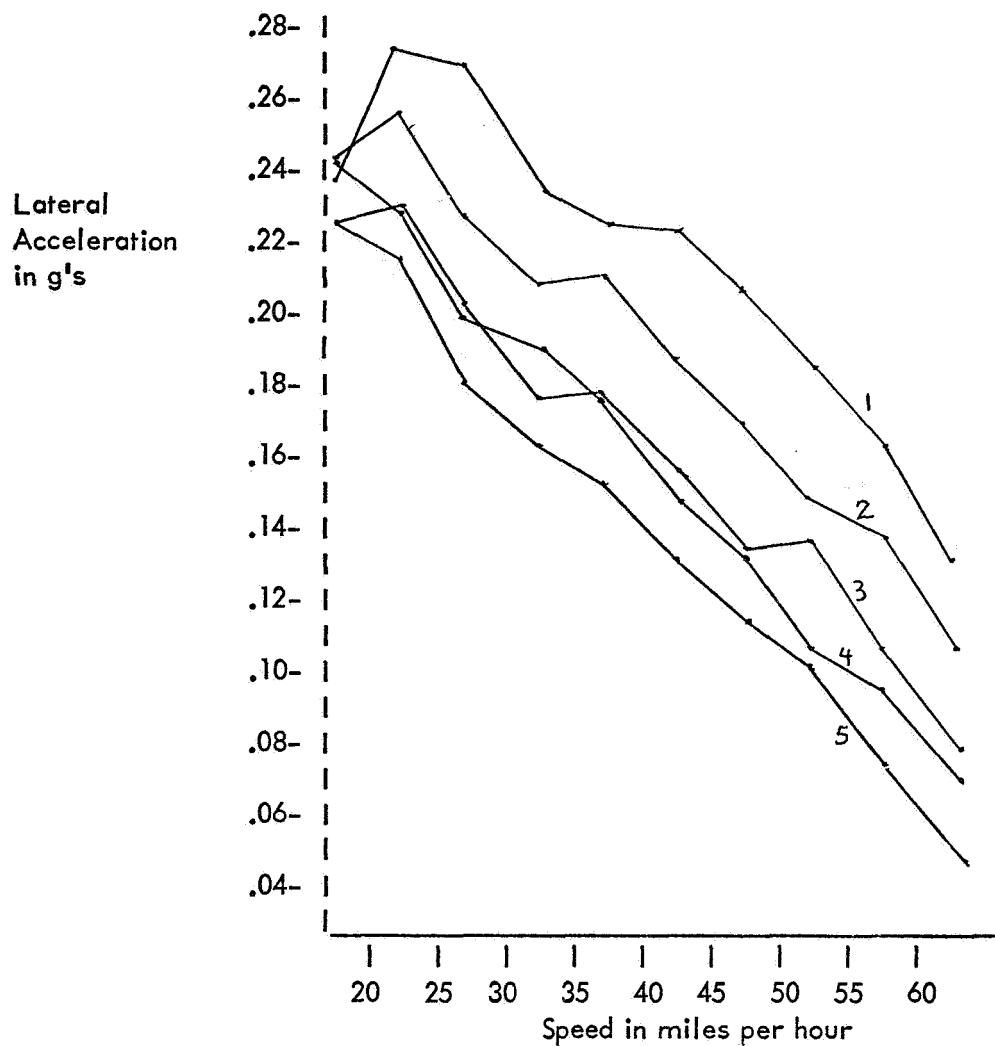


Figure 2.—Lateral acceleration as a function of speed in curves for subjects grouped according to overall speed.

TABLE 1.—SPEED AND ACCELERATION SCORES IN CURVES AS A FUNCTION OF ADVISORY SIGN SPEED VALUES

Sign	15	20	25	30	35	40	45	50
\bar{V}	21.9	30.6	32.1	37.5	41.9	41.3	44.6	50.8
Sd	3.6	4.8	4.5	5.2	5.6	5.3	6.1	6.4
\bar{A}	.22	.25	.23	.22	.21	.19	.16	.15
Sd	.10	.07	.08	.06	.07	.07	.07	.05
n	186	281	231	232	631	380	465	227

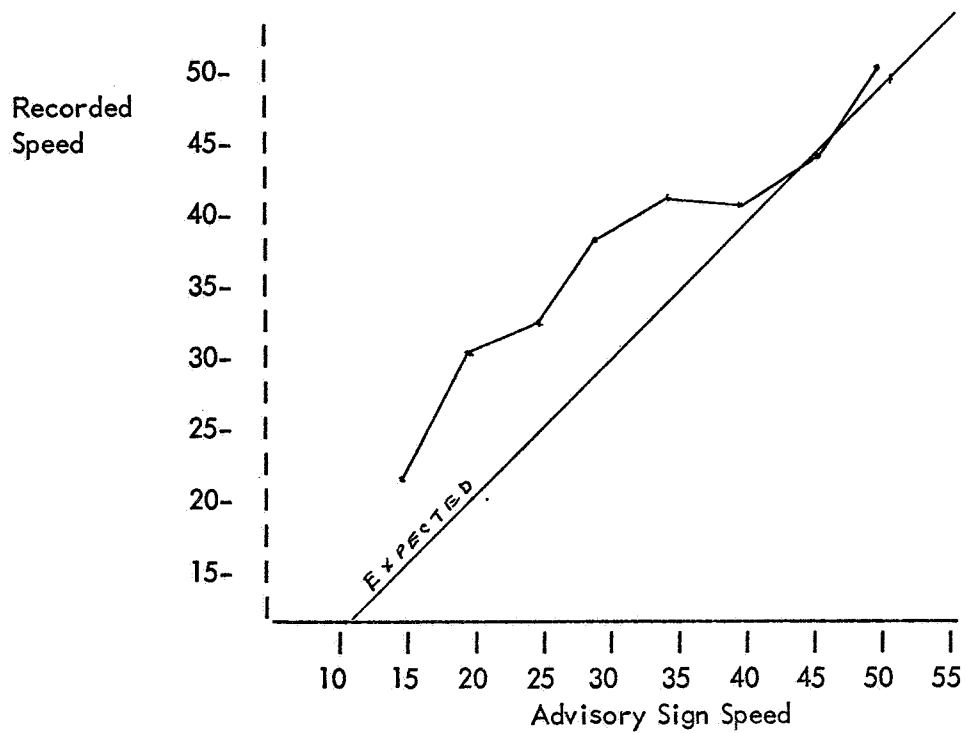


Figure 3.—Recorded speed in curves versus the posted advisory speed for the curves.

REFERENCE

1. Ritchie, M. L., McCoy, W. K.; and Welde, W. L.: A Study of the Relation Between Forward Velocity and Lateral Acceleration in Curves During Normal Driving. Human Factors J., in press.

29. Studies in Driver-Aided Car Following *

*Robert E. Fenton and Charles L. Shaffer
The Ohio State University*

Two studies, involving a kinesthetic-tactile driver aid built into the head of a control stick, were conducted to investigate driver-aided car following. Two stick configurations, one "front-" and one "side-mounted," were examined, and tests were conducted on both a well-traveled rural road and an unopened super-highway. The results of these tests and a discussion of the problems involved in stick control of an automobile are contained in this paper.

INTRODUCTION

An examination of traffic conditions today—congested roadways, a large number of accidents and fatalities, and extremely powerful automobiles—indicates the need for improvement in our highway system. One promising approach to the dual problems of increasing both highway capacity and highway safety involves highway automation. One automated system that has been proposed (ref. 1) as a solution to these problems would function as follows:

A traveler would manually drive his vehicle to the entrance ramp of an automated expressway, where vehicle control would be turned over to an automatic system. The vehicle would be guided into the expressway traffic stream and remain under automatic control until it arrived at a preselected exit. At this point the driver would be given control, and he would manually drive to his destination.

It is estimated that the use of such a system would increase highway capacity as much as 800 to 1000 percent at speeds of 70 mph. Also, since an electronic control system would be more consistent and reliable than a human operator and have a shorter reaction time, the number of accidents should be reduced.

Such a system must include a manual mode so that a driver can control his vehicle on nonautomated roads and regain control, when necessary, on automated ones. The performance of the manually controlled system must be compatible with that of the automatically controlled one, so that the former would cause no fluctuations in an otherwise automatically

*This research was sponsored by the Ohio Department of Highways, Columbus, Ohio, in cooperation with the Bureau of Public Roads.

controlled traffic stream. The present driver-vehicle system is not satisfactory in this respect as considerable performance variability, both within and among drivers, has been observed even in steady-state situations. In addition, conventional automobile controls (steering wheel and brake and accelerator pedals) are not satisfactory for use with an automatic system and must eventually be replaced by a more compatible type of manual control.

The use of a single manual control unit, such as a control stick, appears promising because the effective human reaction time in braking can be halved with such a unit (refs. 2 and 3). Further, it is relatively easy to obtain compatibility between the automatic mode and a control stick. However, the driver's performance variability is primarily due to his limited information-sensing capability rather than to the type of control. This limitation can be largely overcome by presenting him with additional information. It is especially important that the performance of the driver-vehicle system approximate that of the automatic system in steady-state car following. It is not difficult to design an automatic system that has good performance (ref. 1); however, several investigators (refs. 4, 5, and 6) have reported that a driver is unable to maintain a constant, or even a near-constant, headway under steady-state car-following conditions. It appears that the headway variation is due to a driver's inability to detect slow relative motion between himself and a lead car; thus, this variation can be reduced by presenting him with additional information. Bierley (ref. 5) obtained such a reduction by using a galvanometer display which indicated either an instantaneous deviation from a desired headway or a linear combination of headway deviation and relative speed. In a similar study, Gantzer and Rockwell (ref. 7) found that the use of a discrete visual display greatly reduced headway and velocity variances at large headways (200 feet); however, little improvement was observed at small headways (70 feet). Their subjects were apparently too concerned with watching the lead car at the short headway to pay attention to the display.

Fenton and Montano (ref. 6) studied a kinesthetic-tactile display wherein headway information was imparted to a subject via the position of a "finger" mounted in the head of a control stick. This stick was used to replace conventional automobile controls. Sizable reductions in both headway variance and relative velocity variance were obtained for headways from 23 to 60 feet; hence, the use of this control-display combination for improved headway control looks promising.

The research discussed here is an extension of this study and is part of a continuing effort to develop a manual mode which is compatible with an automatic vehicle control system currently being developed (ref. 1). It was desired to determine how effectively a driver would adapt to stick control in real-world car following and also to further investigate the use of a kinesthetic-tactile display in car following.

TEST VEHICLE INSTRUMENTATION

A fully rotational control stick was installed in a 1965 Plymouth sedan to replace the steering wheel, accelerator pedal and brake pedal. The first stick tested was located in front of the driver, and displacement from the neutral position was proportional to the applied force (the second stick tested is described under "An Improved Control Device"). The vehicle was steered by moving the control stick head to the left or right, accelerated by moving forward, and braked by pulling it back.

A kinesthetic-tactile display was built into the head of the control stick. Enlarged views of the head are shown in figures 1 and 2, the "finger" protruding in the first case and recessed in the latter. A driver determines the magnitude and direction of an error by feeling the finger displacement from the neutral or "flush" position. For example, if the

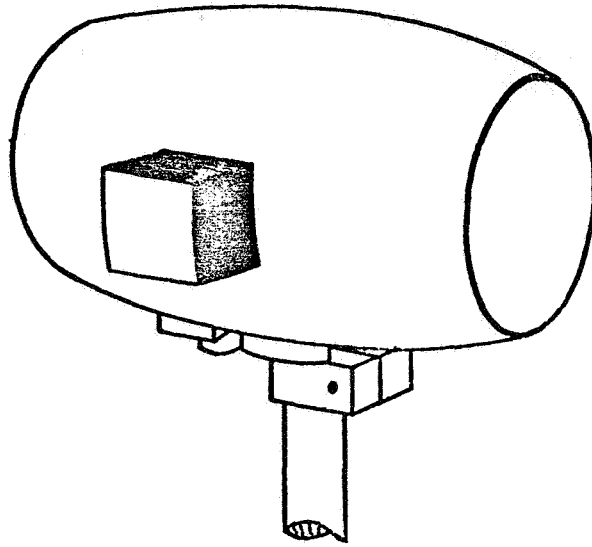


Figure 1. —Head of the control stick with finger protruding.

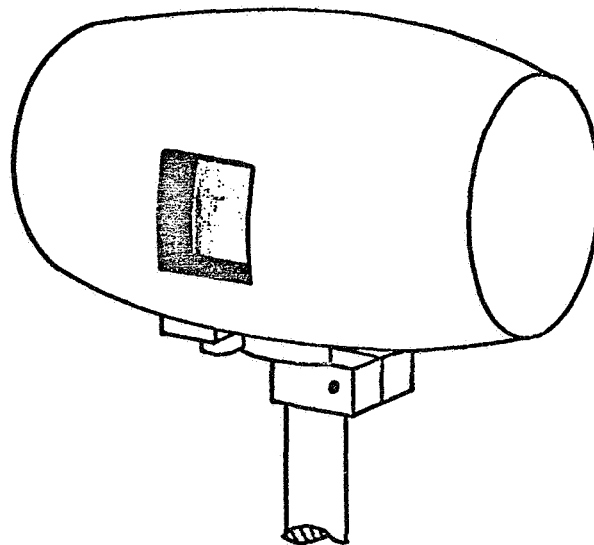


Figure 2. —Head of the control stick with finger recessed.

headway decreases so that it is less than some desired value, the finger recesses in proportion to the deviation from this threshold. A subject counteracts this decrease by pulling back to decelerate; thus, he "follows" the finger.

The position of the finger is controlled by a servomechanism with a speed of response less than 0.4 second. The servo input signal (x) is a linear combination of the relative velocity (v) and the headway deviation (Δh) between the test vehicle and a lead car

$$x = (k_1 v + k_2 \Delta h) k \quad (1)$$

Here, k_1 and k_2 are constants, and k is a gain term. The constant k_1 determines the finger-position sensitivity which refers to finger movement per unit headway deviation. For example, a value of $k_1=0.064$ in./ft corresponds to maximum finger displacement of 0.7 inch for a headway deviation of 11 feet. The mathematical ratio $\left(\frac{k_2}{k_1}\right)$ defines the term aiding ratio, and it provides a relative measure of how much derivative information is supplied to a subject.

The instantaneous headway deviation and relative velocity were measured by means of a mechanical takeup reel or "yo-yo" which was a modification of the one developed at General Motors (ref. 5).

PREVIOUS EXPERIMENTS

The first set of experiments were conducted during the fall of 1966 and have been discussed elsewhere (ref. 6). However, since it is desired to use some of these results, a summary of part of this earlier study is presented here.

The experiments were conducted on a partially blacktopped country road which was 16 feet wide and very rough. The road was open to traffic, and hence the subject-drivers had to watch out for other vehicles. The car-following performance measures were headway variance and relative velocity variance.

A pilot study was first conducted to select satisfactory values for k_1 and k_2 in equation (1). Suitable values were found to be

$$k_1 = 0.064 \text{ in./ft}$$

and

$$\frac{k_1}{k_2} \approx 1.0$$

The primary tests were conducted at average lead-car speeds of 30 and 40 mph. The subjects were instructed to maintain a constant headway equal to 1 ft/mph of average speed, e.g., 30 feet at 30 mph. Headway was defined as the distance from the subject to the rear bumper of the lead vehicle. The actual intervehicular spacings differed from the headway by approximately 7 feet. The tests were conducted for both aided and unaided cases. In the latter, the finger was locked flush with the head of the control stick, and the subject received no displayed information. One obtains essentially the same car-following performance with stick control as with conventional control under this condition (ref. 8); hence, the variance values obtained from the unaided stick-driving situation should approximate those which would have been obtained if conventional controls were used. Thus, these values can be used as a reference against which the aided driving data may be compared.

The results obtained from two subjects are plotted versus lead-car average speed in figure 3. Note the large decrease in headway variance—an average of 95 percent—for the aided over the unaided case. The corresponding relative velocity variances are shown in figure 4. Here, one can observe that reductions of some 75 percent were obtained using the display. It should be noted that similar reductions in both headway variance and relative velocity variance were also obtained for other subjects under corresponding conditions.

Although excellent car-following performance was obtained using this control-stick-display combination, it was not satisfactory for other reasons. First, there was no "feel" to the brakes or accelerator—a condition which often resulted in a jerky vehicle ride. Second, the steering was generally satisfactory; however, some cross-coupling between longitudinal

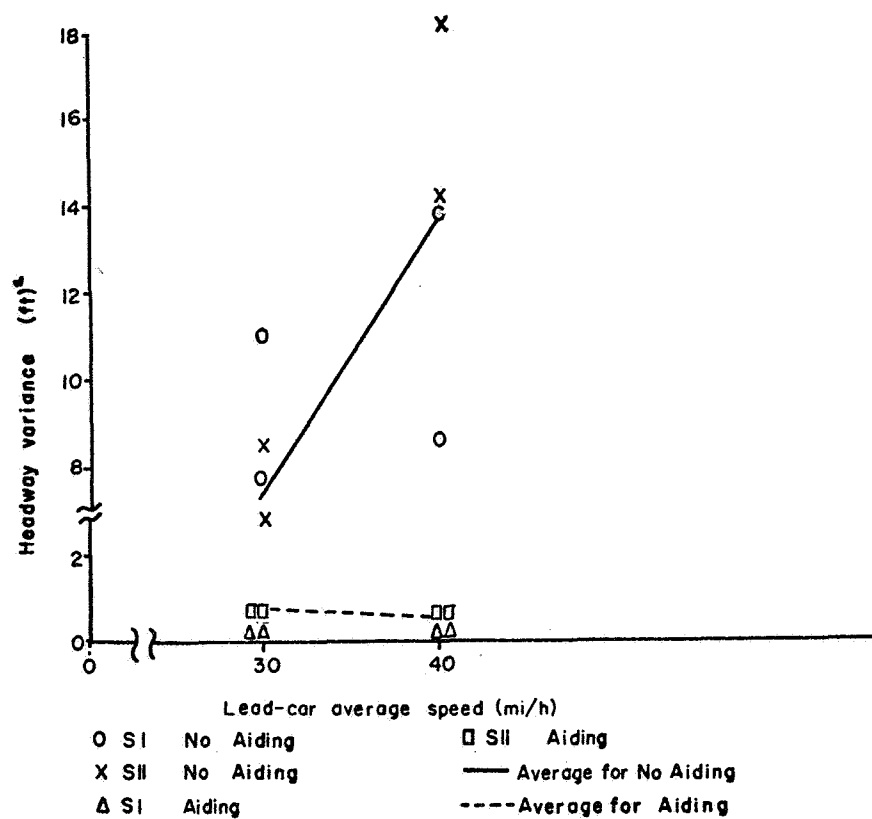


Figure 3.—Headway variances for front-mounted control stick.

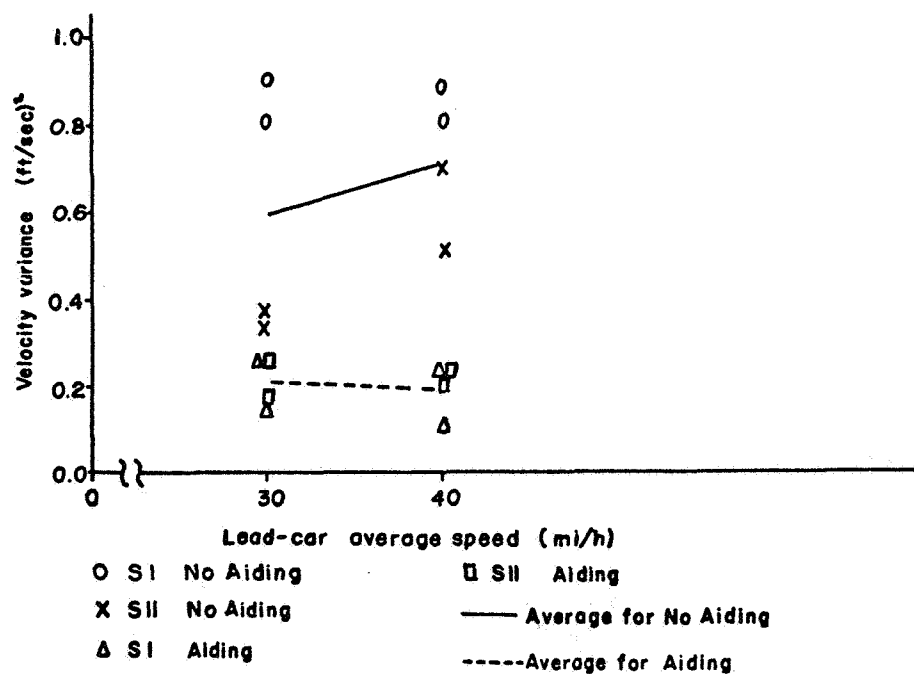


Figure 4.—Relative velocity variances for front-mounted control stick.

and lateral control motions did exist because of the fully rotational characteristic of this control stick. Finally, the subjects were not physically comfortable when driving.

AN IMPROVED CONTROL DEVICE

In light of the deficiencies of the first control stick, it was decided to conduct the remaining experiments with the improved unit shown in figure 5. It was mounted between the driver and the passenger and had a built-in forearm rest to relieve arm fatigue. The cross-coupling problem was minimized by mounting the stick on a carriage which can only be moved in the longitudinal direction, and by providing for only lateral motion of the stick head as indicated in figure 5. The control operations were effectively decoupled when this unit was used.

Synthetic feel was provided by coupling a small hysteresis motor to the carriage. This produced an output, proportional to the velocity of the control movement, which tended to retard the control movement. This velocity damping resulted in a smooth, comfortable ride. The control stick head-display combination previously described was used with this improved control device.

Some initial tests were conducted on the country road to determine if these control-device changes resulted in any significant differences in a subject's car-following performance while using the tactile display.

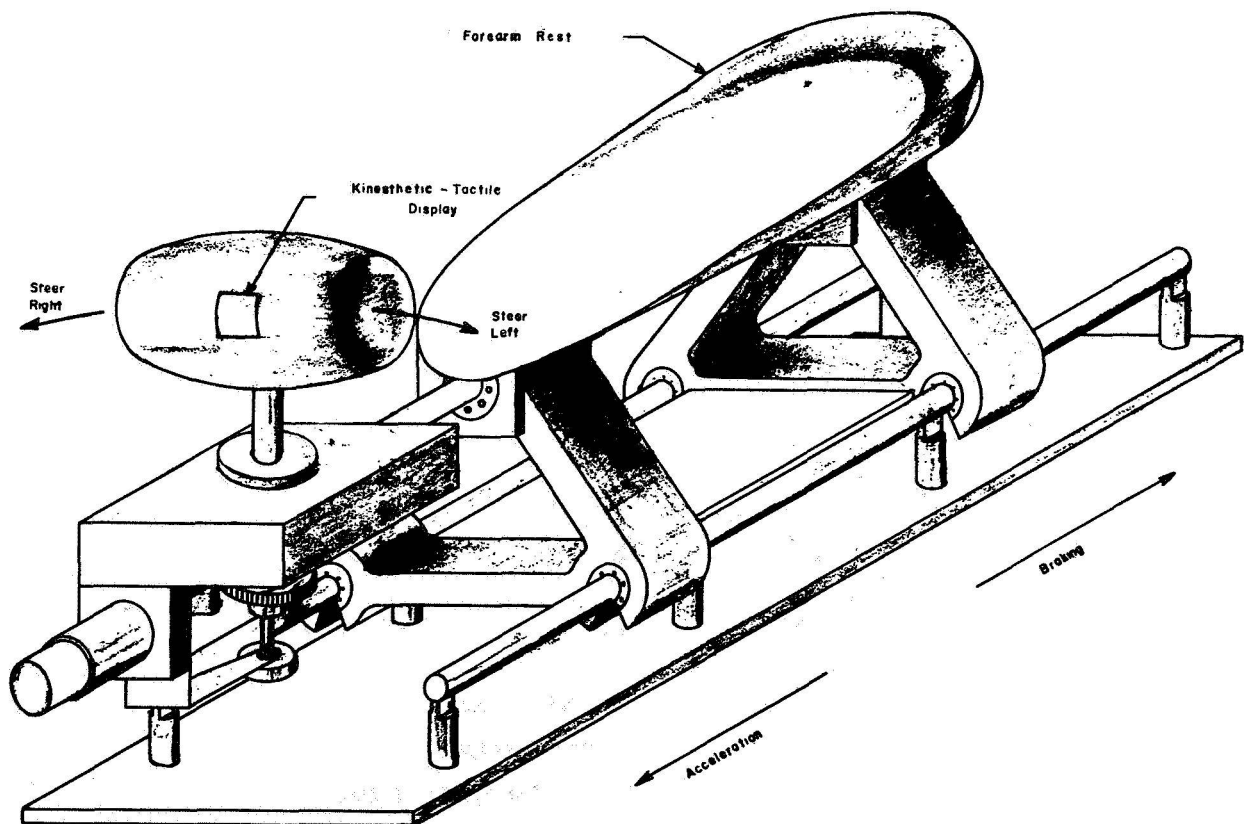


Figure 5. —Side-mounted control stick.

Some results obtained from car-following tests for both the original and the modified control sticks are shown in figures 6 and 7. Here averaged data—headway variance and relative velocity variance—for three subjects have been plotted versus aiding ratio for both control devices. Note from figure 6 that essentially the same headway variances were obtained using the two controls; however, a marked performance difference was obtained for one subject. The same phenomenon is apparent for the relative velocity variances. It is not possible to draw any firm conclusions from these data, since the original data were collected from experienced subjects in 1966, and the other data from the same experienced, but unpracticed, subjects nearly a year later.

It should be noted that the subjects uniformly preferred the improved control and said that more stable two-dimensional driving was obtained with the improved device. Also note that car-following performance was basically comparable for the two control devices. However, the subjects said that an aiding ratio of 1.5 was preferable to the value of one, which had proven satisfactory for the earlier tests.

THE "IDEAL" ROAD EXPERIMENTS

The test site for all remaining experiments was an unopened section of interstate highway which closely approximated an ideally smooth, straight and level roadway. This was in contrast to the rural road previously described. Two experiments were conducted for the following reasons:

- (1) To substantiate earlier results but under different conditions including changed control, different subjects, and a more ideal road.
- (2) To investigate asymptotic stability in driver-aided car following.

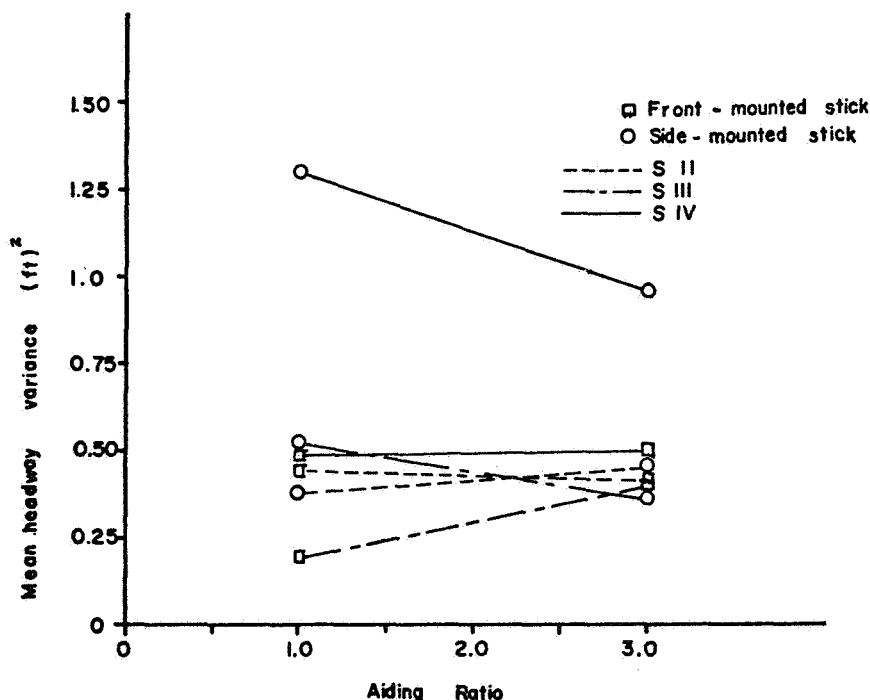


Figure 6.—Headway variances for two control stick types.

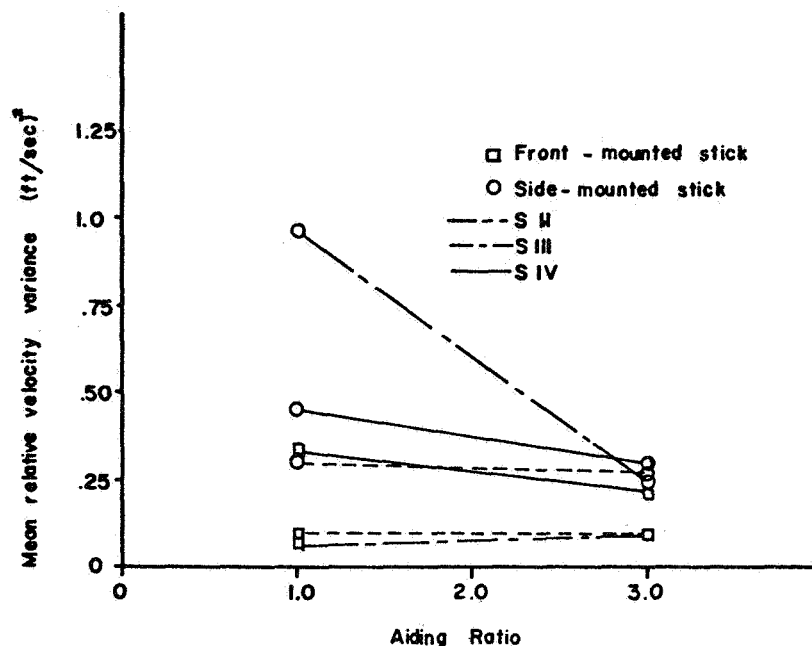


Figure 7.—Relative velocity variances for two control stick types.

PERFORMANCE EXPERIMENT.—The basic experiment was a 4 by 4 full-factorial design where the factors were subject and aiding ratio. There were four replications over trials.

The subjects used in this study were four male employees of the laboratory ranging in age from 20 to 26 years. These subjects are hereafter identified as subjects II, IV, V, and VI. Two of these, II and IV, had participated in the experiments previously described. All subjects were given extensive practice in car following, both with and without the tactile display. At the end of practice, their performance had approached an asymptotic value.

The aiding ratios used were ND, 0.5, 1.5, and 3.0, where ND corresponds to the no-display case. Here, as before, the finger was locked flush with the control-stick head for this case.

In a typical experimental session, the subject followed the lead car, a 1965 Plymouth station wagon, whose speed varied between 37 and 43 mph. He was told to maintain a constant headway of 40 feet between the lead car and the test vehicle. He drove back and forth on a 4-mile length of highway three times, this corresponding to six trials. The first two were practice trials, thus allowing the subject to accustom himself to the condition of aiding employed. Headway and velocity data were collected during the last four trials. The order in which aiding ratios were presented to subjects was counterbalanced to control learning order effects. It should be noted that subjects were given an auditory-paced inspection task to perform as they drove to increase the sensitivity of the performance measures.

An ANOVA was performed on the data obtained in this experiment. Here subject and aiding ratio were considered as random factors, and trials within subjects were taken as the within-cell source of variance. An ANOVA summary table for the headway variances is given in table 1. Note that aiding ratio was the only significant source of variation at the 0.05 level. The mean headway variances obtained from each subject are plotted versus aiding ratio in figure 8. Note that substantial reductions in variance were obtained for the aided relative to the unaided case.

TABLE 1.—ANOVA SUMMARY TABLE FOR HEADWAY VARIANCES

Source	Symbol	Sum of Squares	df	Mean Squares	F
Subject	A	3.21304	3	1.0710	0.8676
Aiding Ratio	B	32.97927	3	10.9931	8.9049*
	AB	11.11049	9	1.2345	1.9340
Within cell	-	30.63840	48	0.6383	-

*Significant at 0.05 level.

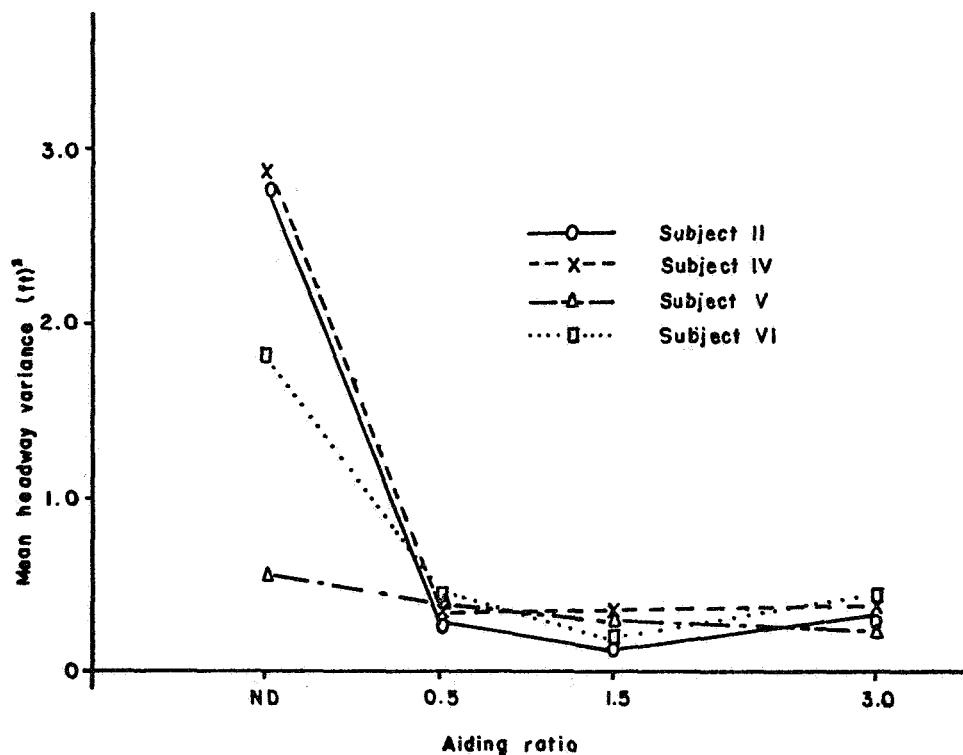


Figure 8.—Individual headway variances (within subject differences averaged out).

An ANOVA summary table for the relative velocity variances is contained in table 2, where it can be observed that aiding ratio and subject x aiding ratio were significant sources of variation. The reason for this significant interaction is clearly evident in figure 9. It is interesting to note that subject was not a significant factor when the display was used. This is consistent with results obtained from earlier tests (ref. 6).

Thus, in general, it can be said that the results reported here were somewhat the same as those obtained from earlier tests. That is, a substantial improvement in car-following performance was obtained using the tactile display as opposed to the no-aiding situation—even in such widely divergent driving situations as have been used in these studies.

TABLE 2.—ANOVA SUMMARY TABLE FOR RELATIVE VELOCITY VARIANCES

Source	Symbol	Sum of Squares	df	Mean Squares	F
Subject	A	.00328	3	.00109	0.8545
Aiding Ratio	B	.03773	3	.01257	9.8396*
	AB	.01150	9	.00128	7.1878*
Within cell	-	.00853	48	.00018	

* Significant at 0.05 level.

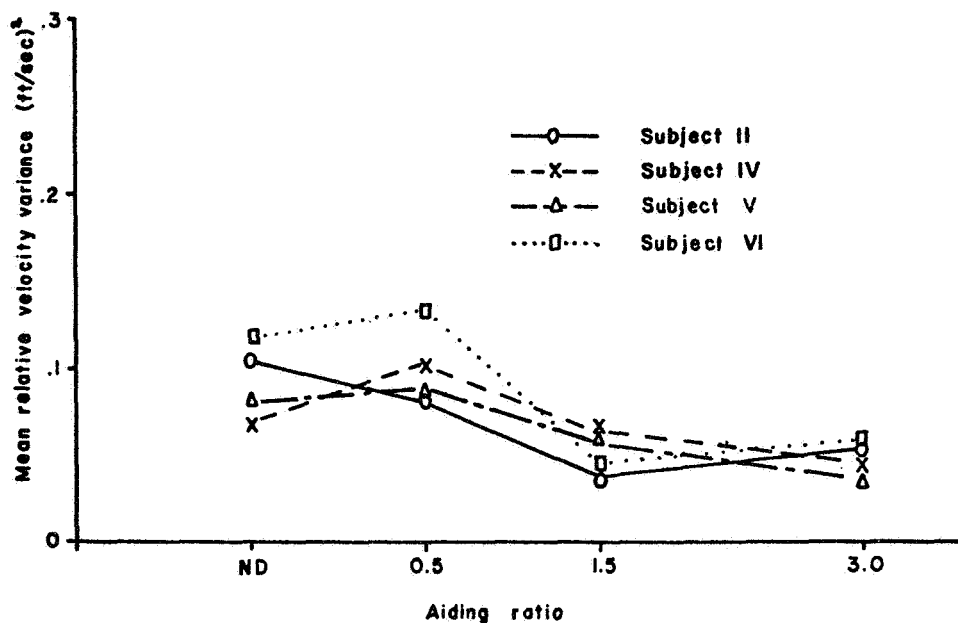


Figure 9.—Individual relative velocity variances (within subject differences averaged out).

The effect of real-world highway conditions upon tracking performance, exclusive of other effects, was obliquely approached during this testing. Unfortunately, only one subject was tested under the same conditions in both these and the earlier front-mounted stick tests, with the exception of type of road and control device. The average performance data for this subject for the no display and "satisfactory" display conditions are plotted in figures 10 and 11. The interesting effect here was that the smooth superhighway and new control condition resulted in much better performance than the rough road and original control, when the subject was tracking the lead car with no display. However, when the display was used, the condition of the road and control appears to have no effect upon performance. Note that no direct statistical comparison of these conditions was possible because of numerous minor differences between the two cases.

ASYMPTOTIC STABILITY STUDIES.—The aided driver-vehicle system can be represented by the block diagram shown in figure 12. It is assumed here that the driver is primarily concerned with longitudinal control and that only small lateral corrections need

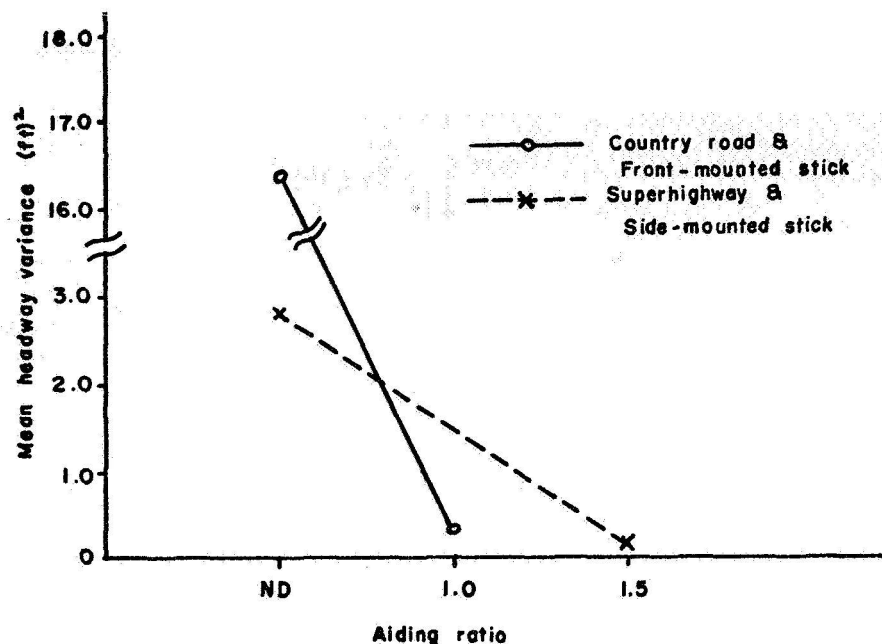


Figure 10.—Headway variances from both front- and side-mounted control sticks (SII).

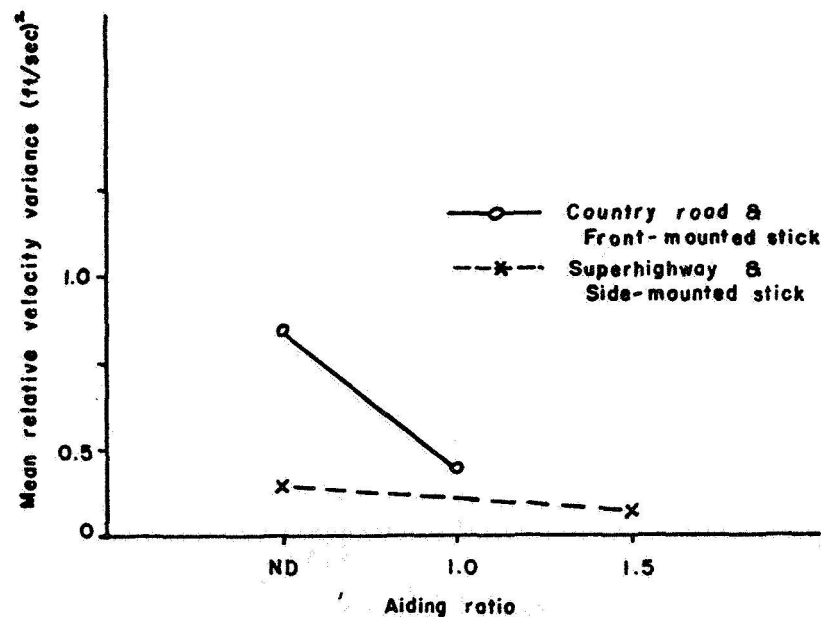


Figure 11.—Relative velocity variances from both front- and side-mounted control sticks (SII).

be made, corresponding to the case of steady-state car following on a long, straight, and uncrowded superhighway (the "ideal" highway used for the experiment described in the last section approximates such a highway). Note the presence of several inputs to the driver—his normal visual input, a display input, and one from acceleration forces. If the driver

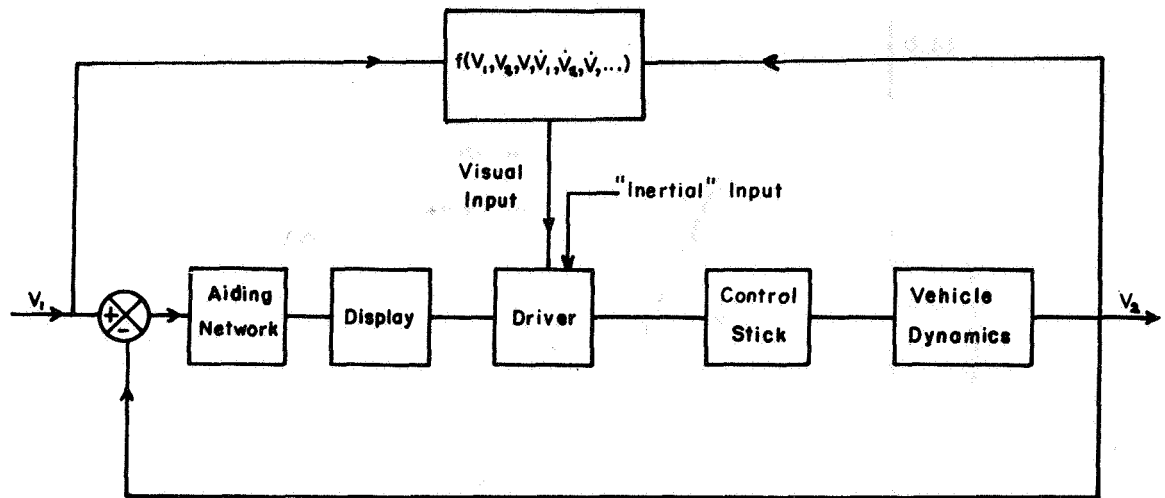


Figure 12. -Block diagram of driver-aided system.

receives sufficient information visually and from the display to control the vehicle, then the system may be treated as having a single input and a single output. A linear model $G(j\omega)$ can be obtained for such a system by using well-known techniques of time-series analysis (ref. 9). This model will be of the form

$$G(j\omega) = \frac{\phi_{v_1 v_2}(j\omega)}{\phi_{v_1 v_1}(j\omega)} \quad (2)$$

where

$\phi_{v_1 v_2}(j\omega)$ = cross-power spectrum of lead and driven car speeds,

and

$\phi_{v_1 v_1}(j\omega)$ = power spectrum of lead-car speed.

These power spectra are readily obtained from certain operations on time records of the lead-car speed (v_1) and the driven car speed (v_2). It should be noted that (2) can also be written as follows:

$$G(j\omega) = \frac{V_2(j\omega)}{V_1(j\omega)} \quad (3)$$

where $V_1(j\omega)$ and $V_2(j\omega)$ are the Fourier transforms of v_1 and v_2 , respectively.

Cosgriff (ref. 10) has shown that a disturbance will be attenuated as it propagates down a line of traffic if

$$|G(j\omega)| \leq 1 \quad \text{for all } \omega \quad (4)$$

Hence, the asymptotic stability of a system can be determined from an examination of its closed-loop frequency response.

The experimental procedure was the same as that previously given except that the data consisted of 300-second magnetic tape recordings of v_1 and v_2 for each subject under driver-aided conditions. These runs were conducted at a lead-car average speed of 40 mph, and the desired bumper-to-bumper spacing was 33 feet. Here again, the subject was told to respond to the display so as to maintain an invariant headway between himself and the lead car, which was a 1965 Plymouth station wagon. The driver of the lead car was told to maintain a constant speed; however, his speed generally varied from 37 to 43 mph over the 5-minute run.

Some typical results are shown on the Bode plot in figure 13 where $20 \log_{10} \left| \frac{v_2}{v_1}(j\omega) \right|$ is plotted versus frequency. Four curves, each corresponding to one run by each subject, are shown. Note the similarity of the responses, a characteristic which is highly desirable for a compatible manual mode. In each case, however, the inequality (4) is not satisfied, and the aided driver-vehicle system is asymptotically unstable. This result is not surprising in view of the small mean-squared tracking errors achieved by the subjects under aided conditions. Thus, one could consider a queue of aided driver-vehicle systems as analogous to a set of mechanically coupled railway cars—that is, a disturbance initiated at the beginning of the train would pass unattenuated through the entire train.

It is clear that optimum tracking in the sense of minimized mean-squared error and asymptotic stability do not occur under identical conditions. It would appear that one must be satisfied with nonoptimum tracking performance, if asymptotic stability is to be achieved. This can probably be realized by compensation of the aided driver-vehicle system, or through the use of nonoptimum parameters in the tracking display. Alternatively, the desired headway could be increased and the display sensitivity decreased.

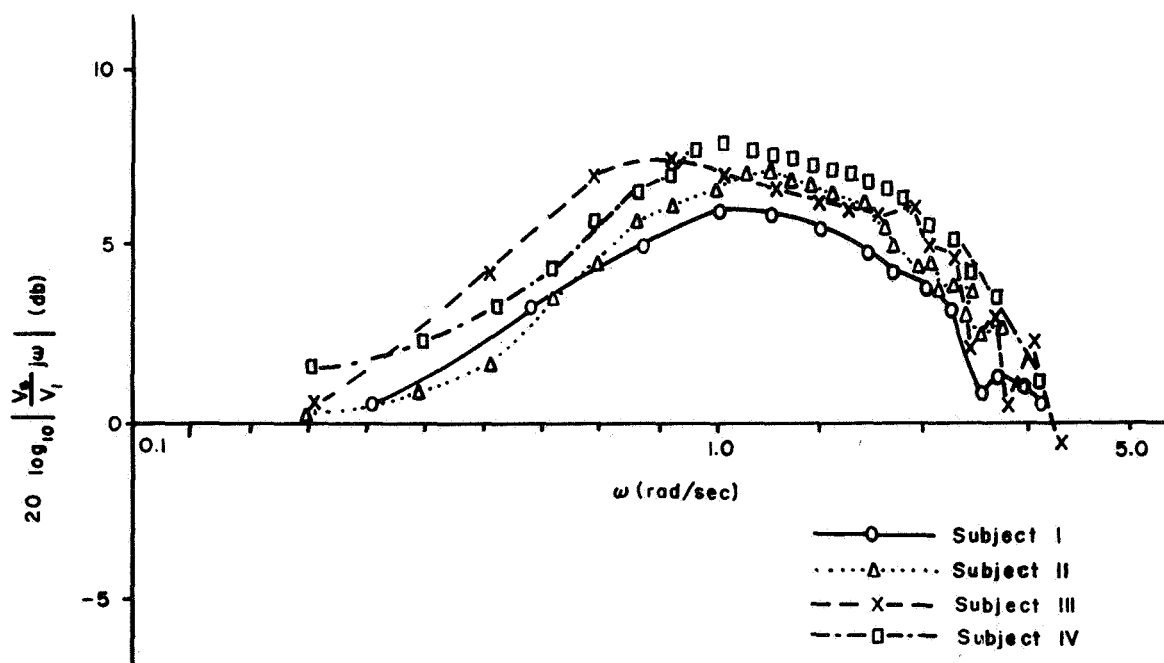


Figure 13.—Describing functions for aided driver-vehicle system.

CONCLUSIONS

The goal of the research reported here is the development of a manual control mode which will be compatible with the automatic vehicle-control system under development at this laboratory. The car-following experiments discussed in this paper were conducted for the following reasons:

(1) To determine how effectively a driver would adapt to stick control in real-world car following.

(2) To further investigate the efficacy of a kinesthetic-tactile display in car following. The subjects easily adapted to both the front- and the side-mounted control sticks. However, the subjects cited two main disadvantages of the fully rotational, front-mounted stick: the lack of "feel," and the cross-coupling between the lateral and longitudinal control motions. Also, the subjects were not physically comfortable when driving.

These problems were largely eliminated by replacing this stick with a side-mounted one in which the longitudinal and lateral motions were largely decoupled. All subjects said that it was quite easy to drive with this second stick, and none had any apparent difficulties during the car-following tests. It is interesting to note that three of the subjects had no difficulty in normal stop-and-go driving or in parking. The fourth subject had no opportunity to drive the vehicle on other than car-following tests.

Excellent tracking performance was obtained using both control-stick-display combinations. Sizable reductions in both headway variance and relative velocity variance were obtained for aided relative to unaided driving using the front-mounted stick on the country road. Reductions in these quantities were also obtained for the aided relative to the unaided case when the side-mounted stick was used on the superhighway. These latter reductions were not as large as the former, since the subjects were able to devote more attention to the lead car when on the empty superhighway.

Excellent tracking performance, however, was not the only criterion of interest, for it was found that optimum tracking and asymptotic stability did not occur under identical conditions. The desirable combination of good tracking performance and asymptotic stability can probably be achieved in a number of ways. Several of these are currently being investigated.

ACKNOWLEDGMENTS

The authors are indebted to several of the laboratory staff for their generously given assistance. In particular, R. Rule assisted in the analysis of the data, and R. Ventola was responsible for much of the control-stick design and construction.

REFERENCES

1. Fenton, R. E.; Cosgriff, R. L.; Olson, K. W.; and Blackwell, L. M.: One Approach to Highway Automation. Proc. IEEE, vol. 56, no. 4, Apr. 1968.
2. Algea, Charles W.: A Model of Driver-Vehicle Systems. In: Study of Electronic Devices as Traffic Aids, Transportation Research Center Rept. No. EES 202-1, Ohio State Univ., Columbus, O., July 1962, pp. 93-105.
3. Konz, S.; and Daccarett, J.: Controls for Automotive Brakes. Paper presented at the 46th Annual Meeting of the Highway Research Board, Washington, D. C.
4. Rockwell, T. H.; and Snider, J. N.: An Investigation of Variability of Driving Performance on the Highway. Systems Research Group Rept. RF 1450, Ohio State Univ., Columbus, O., Sept. 1965.

5. Bierley, Robert L.: Investigation of an Intervehicular Spacing Display. Highway Research Record, no. 25, 1963, pp. 58-75.
6. Fenton, R. E.; and Montano, W. B.: An Intervehicular Spacing Display for Improved Car-Following Performance. To be published in the IEEE Transactions on Man-Machine Systems.
7. Gantzer, D.; and Rockwell, T. H.: The Effects of Discrete Headway and Relative Velocity Information on Car Following. Presented at the 45th Annual Meeting of the Highway Research Board, Washington, D. C.
8. Fenton, Robert E.: Driver-Automobile Interfaces. Communication and Control Systems Laboratory Rept. EES 202A-7, Ohio State Univ., Columbus, O., Sept. 1965.
9. McRuer, D. T.; and Krendel, E. S.: Dynamic Response of Human Operators. WADC Technical Rept. 56-524, Oct. 1957 (AD 110 693).
10. Cosgriff, Robert L.: Dynamics of Automatic Longitudinal Control Systems for Automobiles. In: Theory and Design of Longitudinal Control Systems for Automobiles, Communication and Control Systems Laboratory Rept. EES 202A-8, Ohio State Univ., Columbus, O., Sept. 1965.

30. An Experimental Steering Simulator for Farm Vehicles

C. Saran

University of Puerto Rico

C. W. Suggs

North Carolina State University

Steering of a vehicle can be considered as a problem of manual control systems. With a view to evaluate human performance in the steering of farm vehicles, a steering simulator described in this paper was designed and built. It is the only known attempt in the field. Uniqueness of this simulator lies in that the vehicle itself was disturbed while the target position was fixed. This situation represented a very close simulation of most steered farm vehicles. In most farm vehicles or manually controlled systems the target corresponds to a straight line visualized in a field or some physical object at or near the work area. The disturbances to the vehicle come from the terrain.

With a view to measure human performance in the steering of farm vehicles and to simulate the field surface, the steering simulator discussed in this paper was designed and constructed. In the absence of a widely accepted measure of performance, various response parameters were defined and are being reported by the authors in separate papers (Saran and Suggs, unpublished work).

Research in the area of steering of farm vehicles is almost nonexistent. The nearest attempt in the field was by Wolfe et al. (1955). They described an experimental tractor which dealt with the problem of steering only as one of the many problems of a tractor. It was not designed exclusively for the problem of steering. They had no provisions for simulating the disturbances coming to the steered wheel from the farm terrain.

The simulator discussed in this paper is the only attempt known in dealing exclusively with the steering of farm vehicles. Requirements sought to be satisfied by the simulator are listed below.

- (1) Five driving speeds in the range of 1 to 3.5 mph, typical of farm vehicles in operation, were to be provided.
- (2) A steering carriage simulating the steering mechanism and lateral movement of a vehicle was to be provided.
- (3) A differential gear arrangement for compensatory steering by the driver to counteract the disturbances to the wheel and the carriage was required. Disturbances to the steered wheel are typical of farm vehicles.
- (4) Provisions for disturbing the entire carriage, including the steering mechanism, were to be made.
- (5) Simulation of the disturbances encountered by a farm vehicle when it comes across a furrow wall, stubble or any obstruction which bodily displaces the vehicle in a direction perpendicular to the direction of travel was to be provided.

(6) Sinusoidal disturbances to the steered wheel were to be provided.

With the aforesaid requirements in view, a steering simulator was designed and constructed. It consisted of a frame and drive belt assembly, steering assembly with cross carriage, and instrumentation to measure the inputs and the responses to the man-machine system.

This study is unique in that the vehicle itself is disturbed while the target position is fixed. This situation represents a very close simulation of most steered farm vehicles. In most farm vehicles or controlled systems, the target corresponds to a straight line visualized in a field or some fixed physical object at or near the work area. The disturbances are supplied to the vehicle by the farm terrain.

CONSTRUCTION OF THE SIMULATOR

Moving a vehicle on a stationary ground surface has the same net result as moving the surface and rotating the wheels of the vehicle without moving the vehicle. This latter alternative was the basis of the simulation technique employed. It had the one big advantage of conserving space because it did not require a wide area for the vehicle to move around.

Two 3-foot-long, 6-inch-diameter drums were mounted by means of self-aligning bearings on a rectangular frame made of square steel tubing. A 3-foot-wide rubber impregnated belt was fitted over these two drums. The surface of the belt could be varied to provide for the various farm terrains a vehicle could encounter. The axle shaft of one of the drums was driven by a motor through a chain and sprocket drive. The sprocket ratio could be changed to give various belt speeds. Seat and foot rest supports were mounted on the frame as shown in figures 1 and 2.

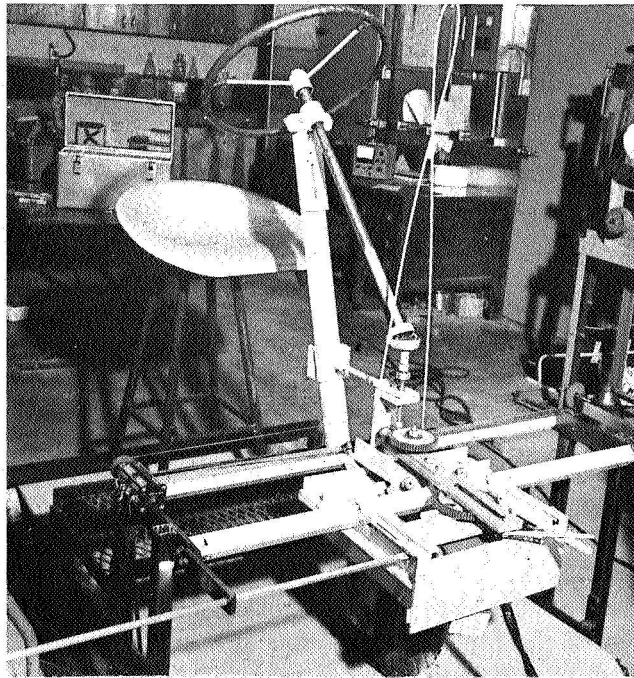


Figure 1.—Steering simulator showing carriage; potentiometer; pulleys and cord to sense carriage location; disturbing bar and differential gear assembly; and steering shaft and supporting frame.

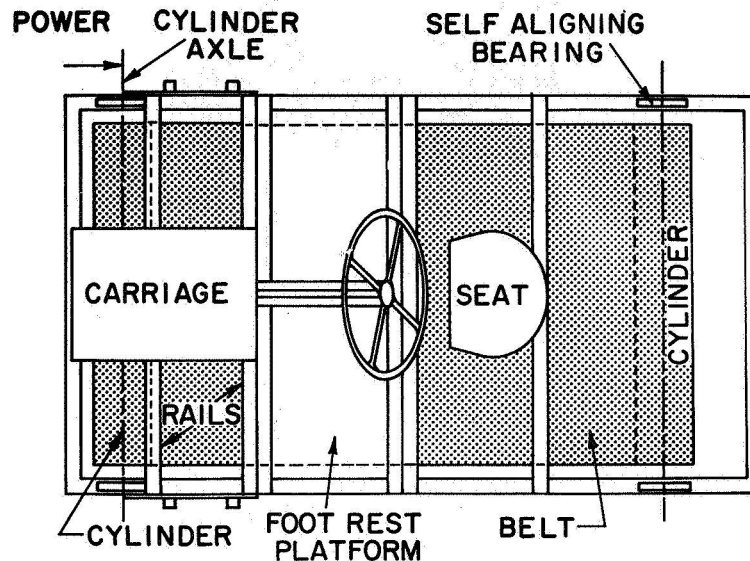


Figure 2. —Schematic diagram of the simulator.

A carriage carrying the steering mechanism and disturbance-providing unit was supported on four 1-inch-diameter cam followers. The carriage could move laterally on two rails made of $1\frac{3}{8}$ by $1\frac{3}{8}$ -inch square section solid steel rods. One of the diagonals of the square cross section was horizontal and along direction of motion of belts. Axle of the cam followers made angle of 45° with the vertical pointing out of the carriage platform. The surface of the cam followers lay flat on one of the surfaces of the rails, (fig. 3). A potentiometer sensed the lateral movement of the carriage. Another potentiometer sensed the movement of the end of the disturbing bar providing the sinusoidal disturbance. This recording was used to obtain the frequency of disturbance.

OPERATION OF THE SIMULATOR

Traction between the ground wheel and the belt moved the carriage laterally on the two crossrails. The motion of the steering wheel was transmitted to the ground wheel through a differential gear arrangement as shown in figure 4.

The steering wheel rotated shaft A. Pinion A on shaft A rotated gear B on shaft B. Gear B and pinion C were rigidly coupled together by a double key bushing and both move freely together on shaft B which is a rigid press fit on to the disturbing bar. Gear D which

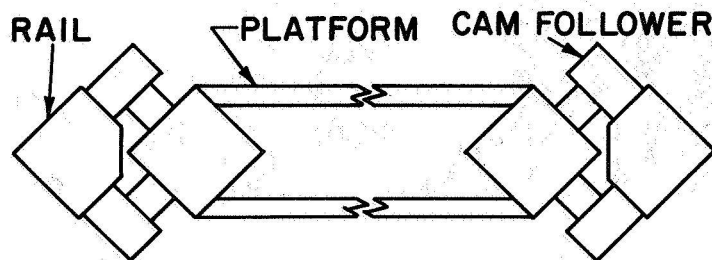


Figure 3. —Sectional view of the carriage platform with rails.

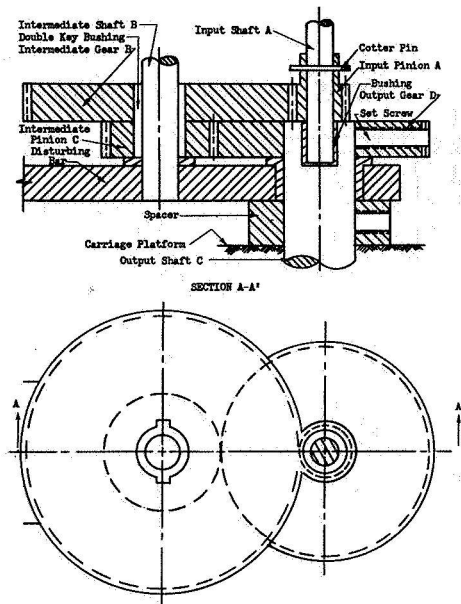


Figure 4. —Differential gear assembly.

is rigidly fixed to shaft C was driven by pinion C. Shaft C supports the ground wheel carriage structure. Rotation of shaft C determines the direction of travel of the ground wheel (figs. 4 and 5).

When the disturbing bar pivoted about shaft C, the cluster of gear B and pinion C rotated around the pinion A. If the pinion A is stationary, this rotation must result in rotation of shaft C. Thus the rotation of pinion A can counteract the rotation of shaft C.

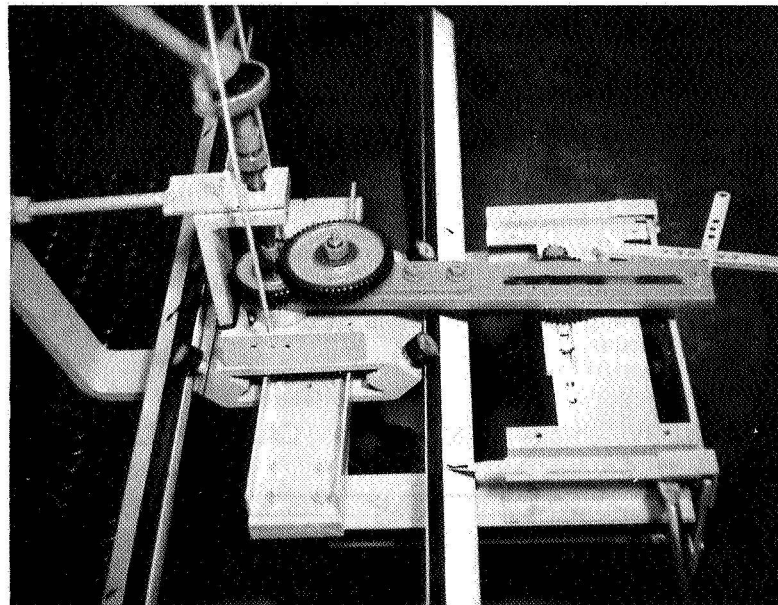


Figure 5. —A close up view of the carriage showing differential gear assembly and the disturbing system.

In order to produce sinusoidal disturbances a 1/2-inch-diameter cam follower was mounted on the flat surface of a gear. This gear rotates around a vertical axis and is driven by an electric motor. The cam follower moves in a slot in the disturbing bar. As the cam follower moves in a circle it causes the disturbing bar to oscillate through an acute angle. The frequency could be changed by changing this gear ratio (fig. 5).

This essentially sinusoidal disturbance simulates steering while traveling along a V-shaped trench or drain. As the vehicle climbs up one wall it stores energy. Then, on its return to the bottom, this energy is converted to kinetic energy. Because of this kinetic energy the vehicle climbs the other wall and the process repeats.

INSTRUMENTATION

The position of the carriage was sensed by a potentiometer. A cord tied to the carriage and passing over the potentiometer pulley at one end and another pulley at the other end rotated the potentiometer shaft. The frequency of oscillation of the disturbing bar was sensed by another potentiometer through bar linkages. The rotation of potentiometers changed voltage signals which were recorded on strip chart recorders.

Although this technique of measuring the exact location of a vehicle at each instant does not measure the "spare mental capacity" (Poulton and Brown, 1961), nor is it similar to the "cone counting technique" (Hoffman and Toubert, 1966), it is much more sensitive to the driver's performance resulting from his ability to maneuver the simulator carriage efficiently.

Heartbeat signals were fed to a Telemedics RKG-100 transmitter from two chest electrodes. These signals were then transmitted to a Telemedics RKG-100 receiver whose output signals were fed to a strip chart recorder. This chart was used as an electrocardiogram.

RESULTS AND COMMENTS

A servo system model of unity feedback servo mechanism was developed and is being reported by Saran and Suggs. Error response to step disturbance was classified into response time, response phase, inactive phase, and correction phase. The transfer function $G(s)$ for the system was

$$G(s) = \frac{C(s)}{D_s(s)} = \frac{1}{(1 + T_1 s) \left(\frac{s^2}{\omega^2} + 2\frac{\rho s}{\omega} + 1 \right)} - 1$$

Saran and Suggs report that with the decrease in belt speed and increase in experience all the time parameters of the response curves increased, and all the following decreased: Percentage of pattern showing overshoot, amount of overshoot per unit disturbance, and area of error in response to sine disturbances. Patterns without overshoot had larger time parameters. Time taken to recover two-thirds of the disturbance decreased with increase in the amplitude of disturbance. Heart rate for the subjects did not show any significant change. There was an interaction of learning with monotony.

ACKNOWLEDGEMENTS

The authors would like to record their deep sense of gratitude to the North Carolina State University at Raleigh for providing the facilities for the work and to the faculty, staff, and the students for their help, guidance, and encouragement.

REFERENCES

1. Brown, I. D.; and Poulton, E. C.: Measuring the Spare Mental Capacity of Car Drivers by a Subsidiary Task. *Ergonomics*, vol. 4, no. 1, pp. 35-40.
2. Hoffman, E. R.; and Toubert, P. N.: The Effect of Changes in Some Vehicle Handling Variables on Driver Steering Performance. *Human Factors*, vol. 8, no. 3 1966, pp. 245-263.
3. Saran, C.: Man-machine Systems Approach to Farm Vehicle Steering Analysis. Ph. D. Thesis, North Carolina State University at Raleigh. University Microfilm, Ann Arbor, Michigan. 1967.
4. Wolfe, J. S.; Bull, D. A.; and Carpenter, L. J.: Influence of Operator's Position on Steering Performance in Row Crop Hoeing. Report No. 41, Nat. Inst. of Agr. Eng., Bedfordshire, Great Britain, 1955.

VIII. DEVELOPMENT OF MODELS BASED ON DISCRETE SIGNAL PROCESSING

31. A Revised Stochastic Sampled Data Model for Eye Tracking Movements*

*L. R. Young, J. D. Forster, and N. Van Houtte
Massachusetts Institute of Technology*

Our sampled data model is revised by changing the pursuit system to be continuous and proportional to target rate. Transient responses of the model are shown to agree in detail with observed classes of eye movements. Target-synchronized and non-synchronized sampler control logic are compared. Predicted latency distributions of a nonsynchronized sampler with unknown intersample time distribution are analyzed, and saccade synchronized input experiments are compared with predicted mean model responses. The results show properties of both types of sampler control logic. The assets, limitations, and extensions of the model are discussed. The sampled data model proves to be a useful tool for predicting eye response based on target movement only.

INTRODUCTION

The discrete nature of the saccadic eye movement control system has been well established (refs. 1, 2, and 3). Several types of discrete systems have been proposed to model eye tracking movements including preprogramed systems, saturating systems, digitally controlled systems, and sampled data systems. It is questionable whether the minute differences between these systems allow differentiation among them in defining, from present research, the exact nature of the eye movement control system. For example, any system at all could be equated to a preprogramed system that yields the observed output to each input. Further, a target-synchronized sampled data system need only choose its stochastic distribution of synchronization times appropriately to model exactly observed latency distributions. The sampled data model, however, is a simple, analyzable, deterministic means of predicting eye movement based on target movement only.

The purpose of this paper, then, is to present our sampled data model, which has been revised in view of recent experiments by several researchers, and to discuss its assets and limitations. Further, the synchronization of sampling instants relative to target motion will be investigated as a possible source of the variability of response observed in following identical experimental stimuli.

*This research was supported by NASA Grant NsG-577.

FURTHER STUDIES

Young (ref. 3) proposed a sampled data model for eye tracking movements in which both the saccadic and pursuit eye movement control system were sampled functions of retinal error. In analyzing model outputs he made the simplifying assumption that the sampler was synchronized to the onset of target motion if the eye had made no saccadic movement in the last 0.2 second. Since Young proposed this model, several studies have shed further light on the eye movement control system.

Rashbass (ref. 4) has shown that the pursuit eye movement control system is independent of the saccadic eye movement control system and sensitive only to target rate rather than target position. Rashbass demonstrated that certain classes of step-ramp target movements generally elicit no saccadic response. Robinson (ref. 2) used these no-saccade step-ramps to show that the pursuit eye movement control system could make changes temporally spaced by the same time as target rate changes for spacings as low as 150, 100, and 75 msec. Robinson also performed variable feedback experiments around only the pursuit system, by eliminating saccadic eye movements from the feedback signal. He observed only smooth continuous changes in pursuit eye movement response. Robinson therefore concluded that while a sampled data system gave good results for the saccadic eye movement control system, the pursuit eye movement control system was continuous rather than sampled. Vossius (ref. 5) proposed a structure of such a system, including a sampled data saccadic eye movement control system and a continuous pursuit eye movement control system, but made no attempt to define operational functions.

Zuber (ref. 6) performed two experiments in attempting to define the nature of sampling in the eye movement control system. His ramp-step-ramp experiment showed that some information available only after the step occurred was used in determining the size of saccadic response to the step. The predictive nature of this result is compatible with non-synchronized sampling and with sampling synchronized a constant or stochastic time after the step. The size of the sampled error depends on the time from the step to the next sampling instant. Zuber's step with variable blank period (no target displayed) experiment showed only random variation of error. This result is incompatible with sampling synchronized a constant time after a target step but is compatible with stochastic target-synchronized sampling and with nonsynchronized sampling.

Beeler (ref. 1) and Wheelless (ref. 7) each noted pulse input results that do not agree completely with Young's model. Beeler noted longer latencies in returning pulses than in continuing pulses for short pulse widths. Wheelless found that in cases where a pulse input elicited no response, the pulse had not gone unnoticed, but in fact delayed response to an immediately following step input in the opposite direction by an average of 40 msec. This result suggests a finite pulse width, with positive and negative portions during some samples canceling each other, forcing a wait for the second sample.

Several other related studies have been carried out, examining the first sampled data model. Jury and Pavlidis (ref. 8) noted that Young's model erroneously predicted pursuit movement to small steps. Johnson (ref. 9) suggested a similar system with two switches. Meiry (ref. 10) and Sugie (ref. 11) presented systems that included the sampled data model as a visual subsystem. Dallos and Jones (ref. 12) and Michael (ref. 13) studied the effects of target movement predictability on the eye movement control system transfer function. Lässig (ref. 14) showed that the transfer behavior of a nonlinear system depends on the incoming signal and that differences in eye tracking transfer behavior with predictable and nonpredictable signals do not necessarily indicate learning properties but may reflect the influence of nonlinearities. Fender presented an excellent layman's review of the area. Milhorn (ref. 15), Milsum (ref. 16), and Jones and Milsum (ref. 17) have reviewed sampled

data modeling of the eye movement control system as an example of modeling physiological systems. Fuchs (ref. 18) performed similar experiments to Young's, using monkeys, and Saslow (ref. 19) investigated target position predictability effects on step input latency.

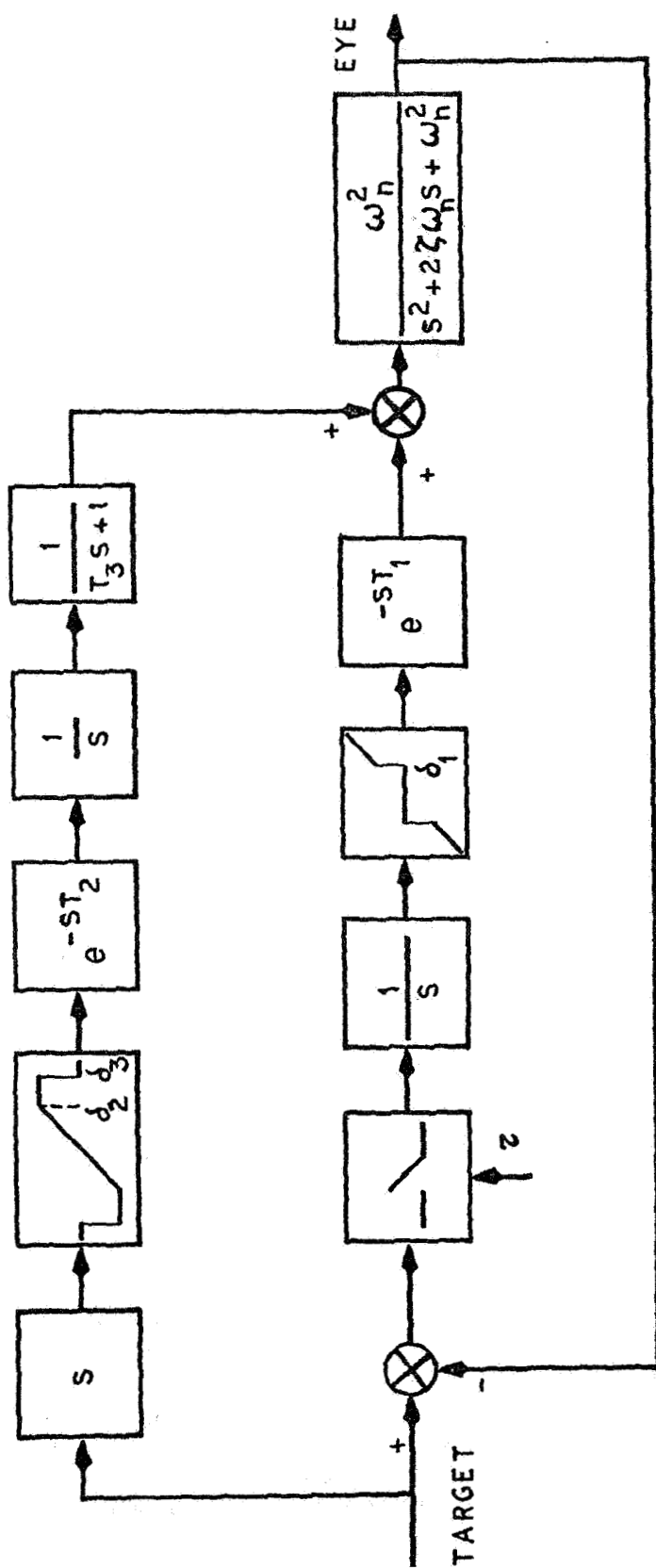
THE REVISED MODEL

Our current model for nonpredictive eye tracking movements is presented in figure 1. The saccadic eye movement control system of Young's original model has only been altered by reducing the delay, retaining the excellent saccadic response characteristics of that model. The pursuit eye movement control system has been made a continuous, open-loop function of target rate. In the pursuit eye movement control system, target position is continuously differentiated, then passed through a nonlinearity that exhibits both saturation and cutoff characteristics. The signal is then integrated to form position commands and delayed before being introduced to the eye dynamics. Eye position is only fed back to the saccadic eye movement control system, so that all position errors, even those due to the pursuit system, are corrected by saccades. The open loop system avoids limit cycles that arise in a continuous closed loop system with delay. This system agrees with Rashbass's findings that the pursuit and saccadic systems are independent and the pursuit system is target-rate sensitive. The system also agrees with Robinson's finding that the pursuit system was continuous and with experimental results that show saccades correcting position errors due to lagging pursuit movements. We now feel that Westheimer's (ref. 20) observation in 1954 that pursuit velocity changes discretely is misleading, since these abrupt changes only seem to appear when facilitated by a saccade.

To predict observed inaccuracies in saccadic eye movements, the slope of the major line in the foveal dead zone nonlinearity must be allowed to be a stochastic variable. A slope less than 1 would produce an undershoot inaccuracy and a slope greater than 1 would produce an overshoot inaccuracy. Similarly, the variability in pursuit movement observed by Robinson may be modeled by stochastically varying the slope of the pursuit nonlinearity. Such a variable slope could account for the ramp responses that Robinson shows have eye velocity exceeding the target velocity, as described in conjunction with figure 2. The definition of these stochastic processes and the possibility of further dynamics in the pursuit system are topics for further research.

Both stochastic target synchronized and nonsynchronized sampler control laws are compatible with Zuber's results and should be further investigated in an attempt to reveal the nature of sampling in the eye movement control system. Target-synchronized sampler control implies that target movement is continuously monitored, and when movement is observed, a stochastic delay elapses before a sample can be affected. The sample is then followed by a constant processing delay before the eye movement response. This theory receives some support from the hypothesis of Robinson (ref. 2) that different mechanisms control fixation and movement in the eye. The fixation system could be continuous with the ability to trigger a sampled data eye movement system. If target synchronized sampling is to match observed pulse input results, a further requirement, that of a minimum time between samples of approximately 200 msec, must be imposed. This target-synchronized system needs only specify stochastic delays before the sample, to be distributed exactly as observed latency minus the constant delay, in order to match exactly step latency distributions of observed eye movement.

Nonsynchronized sampler control logic implies that samples always continue to be taken whether retinal error has occurred or not, and the occurrence of a target movement does not alter the time of the next occurrence of a sample. If this control logic occurs in the eye movement control system, then latency distributions must be explainable on the



SACCADIC DELAY	$T_1 = 150 \text{ ms}$	DEAD ZONE	$\delta_1 = .3 \text{ deg}$	$z = 240 \text{ ms}$
PURSUIT DELAY	$T_2 = 134 \text{ ms}$	SATURATION	$\delta_2 = 30^\circ/\%$	$\zeta = .7$
PURSUIT LAG	$T_3 = 40 \text{ ms}$	CUTOFF	$\delta_3 = 100^\circ/\%$	$\omega_n = 120 \text{ rad/s}$

Figure 1. --Eye tracking movement model.

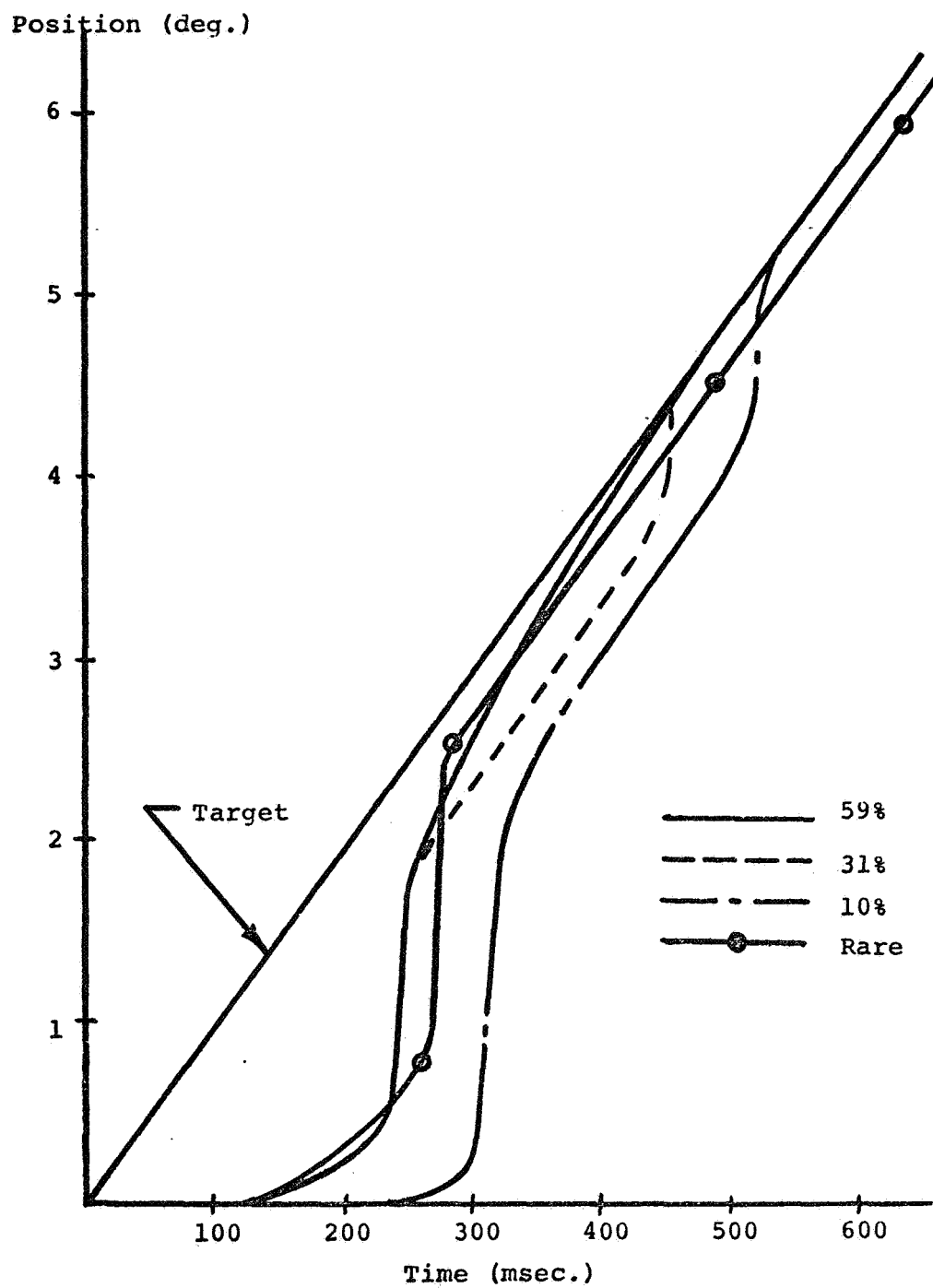


Figure 2.—Eye response to 10 deg/sec ramp stimulus (from Robinson)

basis of the frequency distribution of times between samples. Latency (L) predicted by the model is given by the sum of the time (D) from the occurrence of an input to the occurrence of a sample plus the constant delay (T_1 in fig. 1):

$$L = D + T_1$$

Since T_1 is a constant, it simply translates the distribution of D along the time axis in forming the distribution of L. The shape of the distribution of D, then, must be the same as the distribution of L which should agree with observed latency. The relation of D to the sampling interval τ in which the input occurs is the difference between τ and the time from the sample t as seen in figure 3.

$$D = \tau - t$$

From elementary probability theory, the density function of a random variable that is the difference of two other random variables may be written in terms of the density function of the other random variables as:

$$f_D(D) = \int_{-\infty}^{+\infty} f_t(\tau - D) f_\tau(\tau) d\tau$$

The lower limit of $f_t(\tau - D)f_t(t)$ is zero and the upper limit of $f_\tau(\tau)$ is τ_{\max} . Therefore, the limits of integration reduce to D and τ_{\max} . Since the occurrence of an input does not change the next occurrence of a sample, t may be assumed uniformly distributed over the maximum sampling interval τ_{\max} .

$$f_t(t) = \begin{cases} 1/\tau_{\max} & 0 < t < \tau_{\max} \\ 0 & \text{otherwise} \end{cases}$$

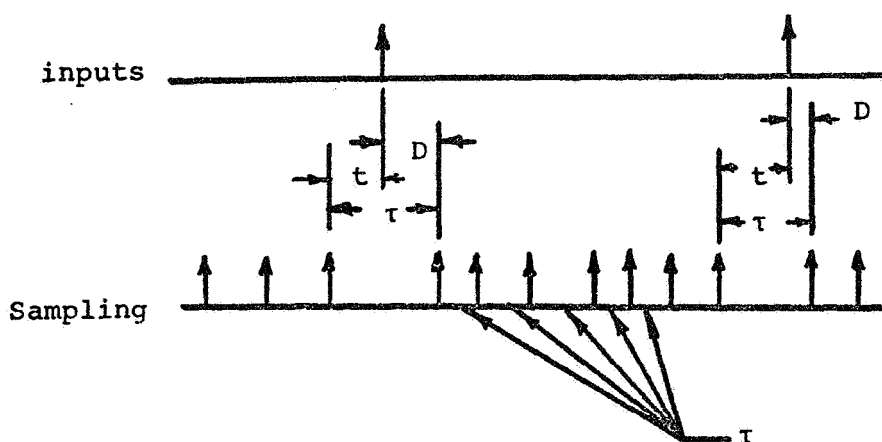


Figure 3.—Nonsynchronized sampler control logic timing.

Therefore $f_t(\tau-D)$ is constant and may be taken out of the integral. Since D is positive by definition the probability density function of D becomes

$$f_D(D) = \begin{cases} 1/\tau_{\max} \int_D^{\tau_{\max}} f_t(\tau) d\tau & 0 < D < \tau_{\max} \\ 0 & \text{otherwise} \end{cases}$$

Since $f_t(\tau)$ must be positive, as D increases less and less positive area will be included in this integral. Therefore $f_D(D)$ must be strictly nonincreasing. This result has been verified by Monte Carlo simulation by Forster (ref. 21).

Clearly an $f_t(\tau)$ could be constructed such that $f_D(D)$ exactly fits the decreasing portion of actual latency distributions. The jump from zero to the maximum value could then be placed to make the integral over $f_D(D)$ unity. Since actual latency distributions do increase, the form of the resulting distribution would be misleading, and so this fitting process loses its meaning. Because of these increases in actual latency distributions for the first approximately 50 msec of possible latency, if nonsynchronized sampling occurs in the eye movement control system, the eye must sometimes be able to shorten the sampling interval when a target movement is observed.

TRANSIENT RESPONSE

With the model as outlined in figure 1, and using the parameters as discussed previously, a number of target-eye movement responses will be discussed.

(1) Figure 4 shows model input-output pairs for pulses, steps, ramps, step ramps, and parabolas. The time scale is superimposed on the input. The pulses indicate the sampling intervals, at a rate of five per second ($\tau=0.2$ second). The timing of each pulse indicates the synchronization of the sample.

Figure 4(a) shows a step input of 5 degrees, with the first sample appearing 100 msec after the input. The saccadic delay of 100 msec adds further latency, such that the response begins 200 msec after the input. The same pattern appears at the return of the input. Figure 4(b) shows three sets of target pulses. Two samples occur during the first pulse, and the output is much longer than the input, returning only after 300 msec beyond the target return, due to the very last processed sample. The second pulse is sampled only once, and the output is of approximately the same duration as the input. Figure 4(b3) shows an input pulse of almost the same size as in the second example, but it has fallen between sampling instants, and leads to no response.

Figure 4(c) shows the model response to ramp inputs of 10 deg/sec. The first example shows that after a pursuit delay of 125 msec, the model is pursuing the target, but with an offset in position. The first sample occurred when the error was less than the 0.5° of threshold of the saccadic system. Following the second sample, a saccade corrects the position error and the eye is on target. At the end of this input, the output of the model exhibits the sawtooth response (ref. 20), with the error corrected after the next sampling instant. The second example shows that in the case of a step-ramp input (the step being in opposite direction of the ramp), there are two saccadic corrections of opposite sign. At the end of the input, we notice an offset error within the dead zone characteristics. The third case uses an identical input to the second one, but a shift in timing of the samples. Unlike the second one, its response does not show any saccades at all.

Figure 4(d) shows the response to a parabola of 10 deg/sec². Notice the continuous pursuit output, interrupted periodically by saccades, which reduce the accumulating position error.

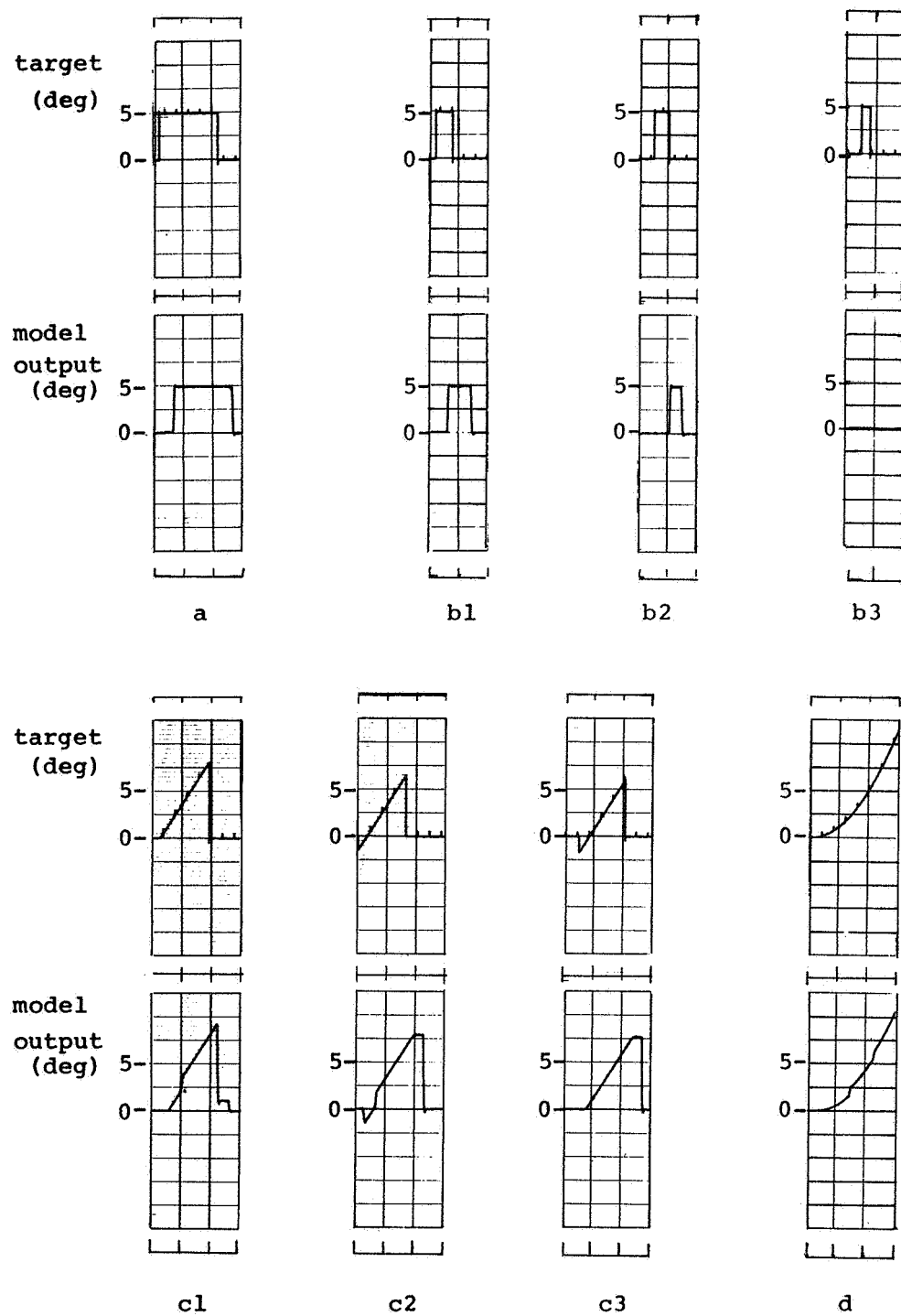


Figure 4.—Model responses to various target inputs: (a) step (5 deg), (b) pulse (5 deg).

(2) As can be seen from figure 4(c), the model predicts different responses to the same type of input. This depends on the synchronization of the first sample with the input. A look at a series of outputs, with varying synchronization, shows the different responses that the model predicts. Figure 4 is a series of responses to a single 10 deg/sec ramp input. The sampling interval of 0.2 second is divided into 10. The first sample is synchronized to the input for the leftmost case, and occurs delayed by an additional 0.02 second for each following case, with the eleventh case a repeat of the first one. There is twice a double saccadic correction, and only once an offset error during the ramp. This variety of response types to a 10 deg/sec ramp agrees with the measured response frequencies presented by Robinson (ref. 2) in his detailed study of pursuit dynamics. For comparison his data is shown in figure 2. At the end of the input in figure 5, we see the probability of having the sawtooth appearance.

Figure 6 shows the series of responses to the step-ramp. Synchronization between first sample and input varies as in figure 5. The successive cases show a variety of saccadic response corrections to this step-ramp input. When the returning target has not moved sufficiently far back toward the origin at the sampling instant, a second saccadic correction in the direction of the ramp is elicited.

Thus, as seen from figures 4(c3), 4(c2), and 6, the model predicts variously none, one, or two saccadic corrections to the same step-ramp input. All first saccades appear within a time interval, which is the sum of τ and T_1 . These responses are in accord with the various pursuit-saccadic categories observed in following a step-ramp by Rashbass (ref. 4) and Robinson (ref. 2). A detailed analysis of the model responses to this case is given in the appendix.

FREQUENCY RESPONSE

A study of the frequency response shows that the model has two modes of tracking. Figure 7 shows the model response to 5 degree input sinusoids at three different frequencies. At low frequencies the model follows the input closely, and saccadic corrections appear at maximum deviation of the target. With increasing frequency, more saccades occur, the output amplitude goes up, and there is an increase in phase lag. It should be emphasized that this is not the predicted eye movement response to a single sinusoid, which involves a predictive mode of tracking. The model frequency response plot, based on single sinusoid analysis, is shown in figure 8. The maximum of the amplitude ratio as well as the phase varies with input amplitude because of the system nonlinearities.

Figure 9 shows the frequency response of the nonpredictive eye-movement system as measured by several investigators (nonpredictive tracking). Although the phase relationships of the model agree well with the experimental, considerable discrepancy is apparent in the amplitude ratio at low frequencies.

SACCADE-SYNCHRONIZED EXPERIMENTS

As seen in the transient response analysis, the synchronization of input to sampling instants is the single most important factor in determining the response of the revised sampled data model for eye tracking movements. Since there is no output at the instant of a sample, the only way to estimate when a sample occurred is by observation of a saccade resulting from a sample. The model predicts that a sample occurred exactly T_1 seconds before every saccade. If an input occurs immediately after a saccade, the sample that caused that saccade and the sample that will cause a response to that input will be separated by only one intersample time. Therefore, if eye movements are controlled by a system such as the revised sampled data model, the distribution of intersample times will be

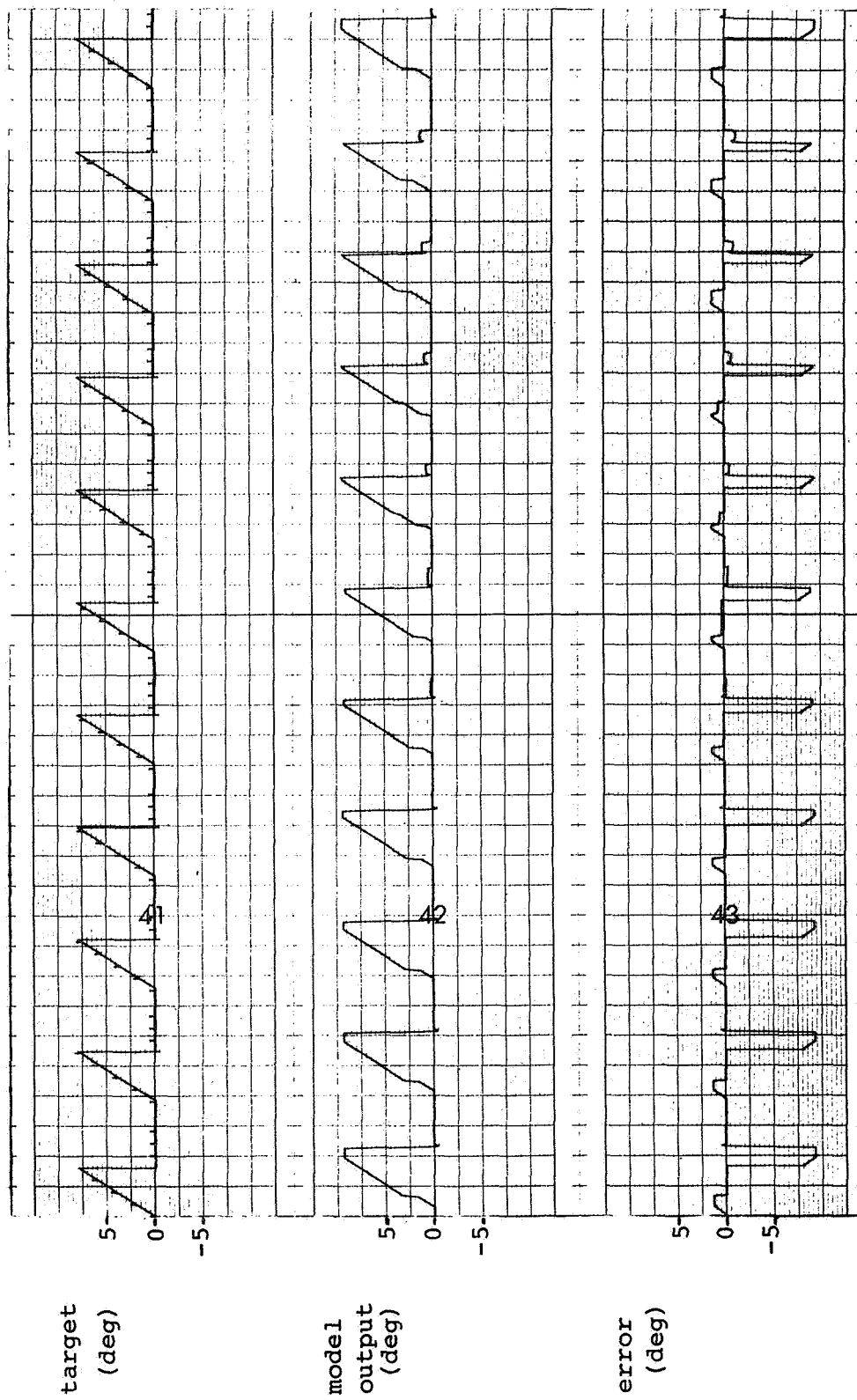


Figure 5.—Series of model responses to a ramp input (10 deg/sec) with synchronization between input and first sample varying from 0.0 to 1.0 τ in steps of 0.1 τ . Plot shows input, output, and error. The time scale is superimposed on the input (time between pulses is $\tau = 200$ msec).

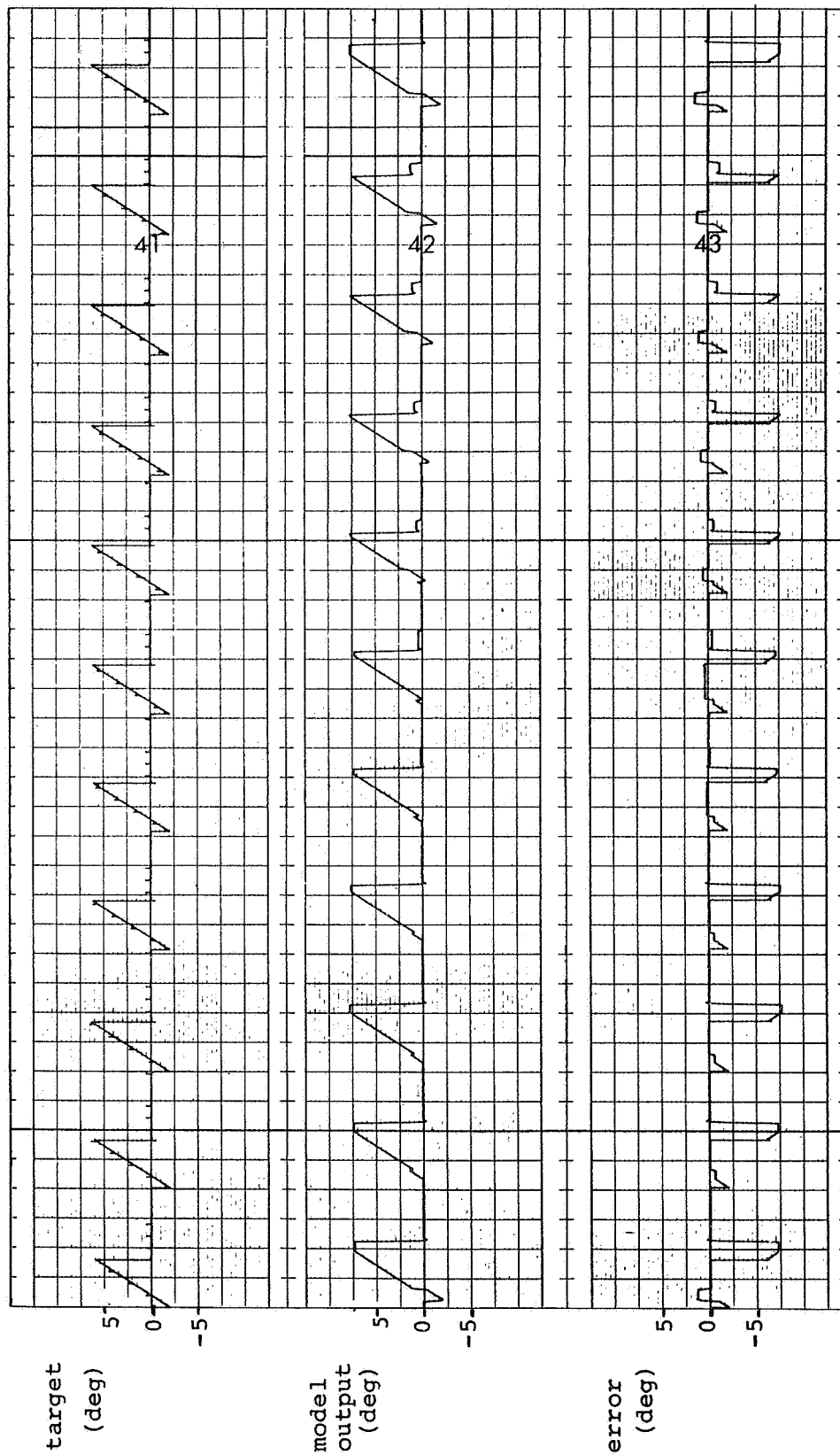


Figure 6.—Series of model responses to a step-ramp input (-2 deg, 10 deg/sec) with synchronization between input and first sample varying from 0.0 to 1.0τ in steps of 0.1τ . Plot shows input, output and error. The time scale is superimposed on the input (time between pulses is $\tau = 200$ msec).

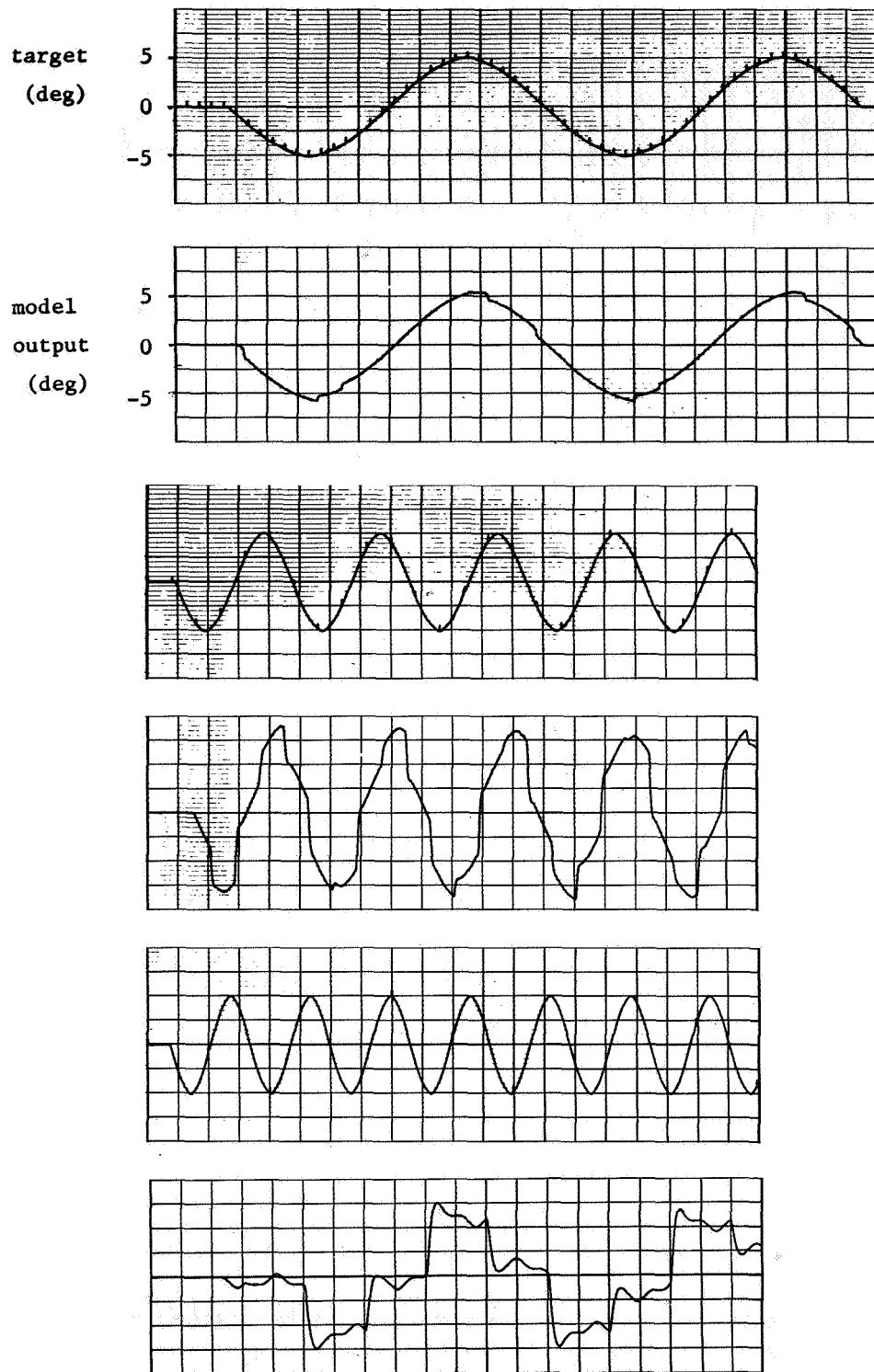


Figure 7.—Model response to a single sinusoid (nonpredictive tracking) for three frequencies: $f=0.194$ hertz (top left), $f=1.04$ hertz (top right), and $f=3.82$ hertz (bottom). The time scale is superimposed on the input (time between pulses is 200 msec).

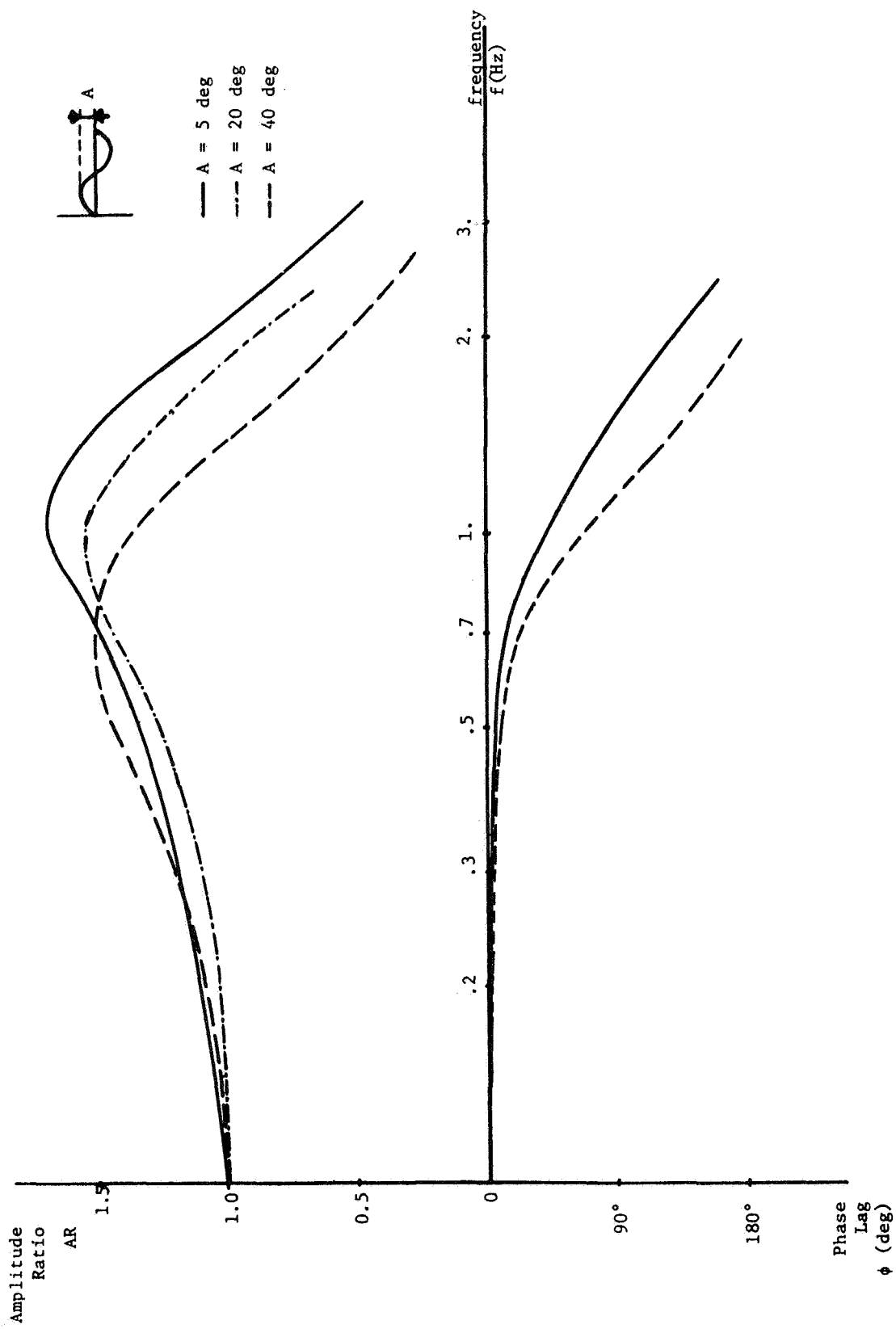


Figure 8.—Frequency response plot of the model for various input amplitudes.

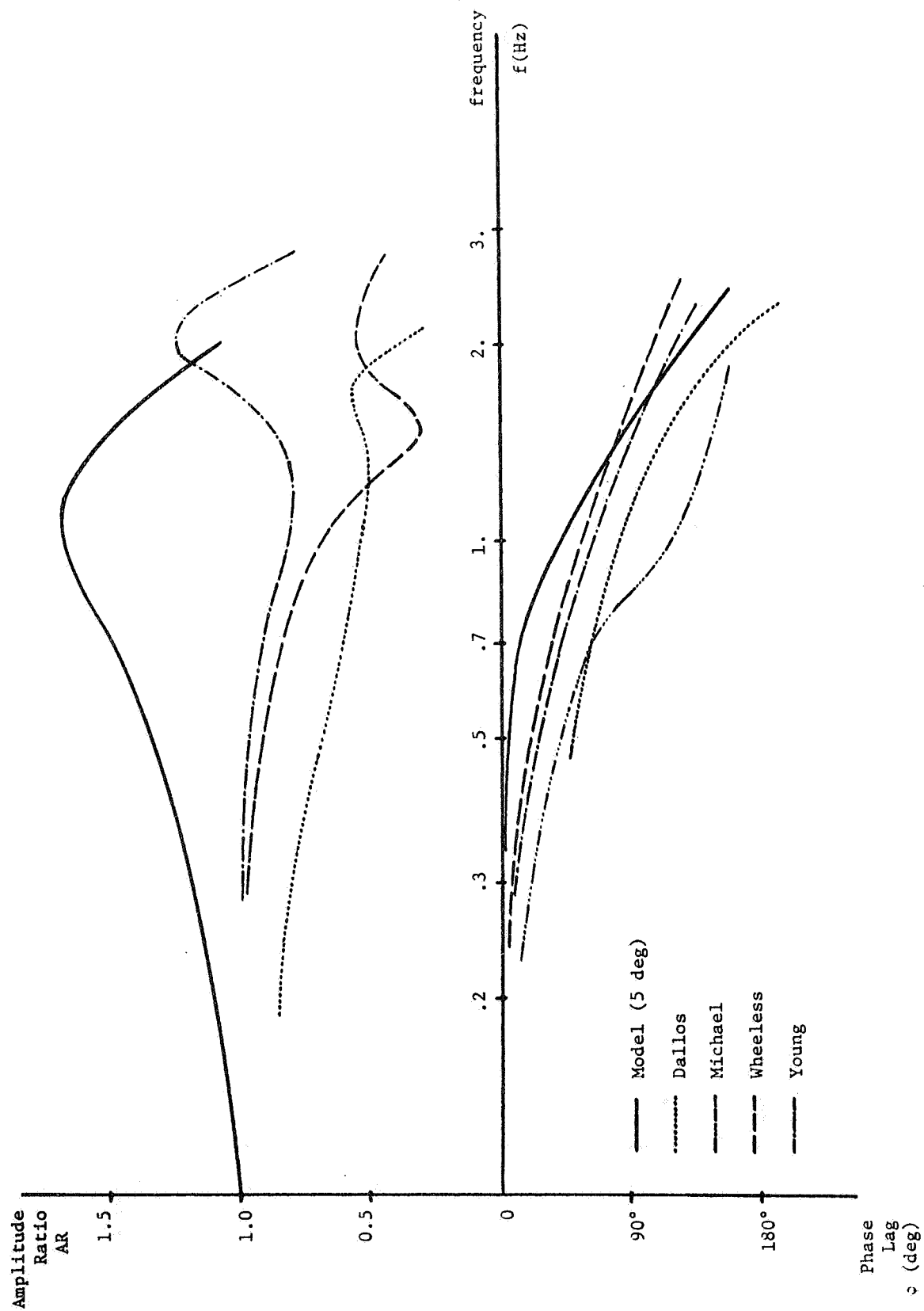


Figure 9.—Frequency response of eye tracking movements (nonpredictive mode) established by several investigators and compared with model responses.

revealed by the distribution of responses to an input that always occurs at the same constant time after a saccade. Further, the sampler control logic of a sampled data model will be revealed by the changes in response distributions as the constant time after a saccade changes.

In a sampled data system with target-synchronized sampler control logic, the average latency to a step will not change as a function of time after the last saccade because each target movement triggers its own sample. If the system also has a minimum time of 200 msec between samples, then for steps introduced a short time after saccades the system must wait for this minimum time to elapse before allowing another stochastic delay preceding a sample. For this sampler control logic, then, as time delay from the last saccade to a step increases from zero, average latency should decrease linearly with slope -1 from an initial high value until the minimum intersample time requirement no longer constrains the system. As delay from saccade to step continues to increase, this system would show no change in latency.

In a sampled data system with nonsynchronized sampling, the distribution of intersample times determines exact latency characteristics. As the time from saccade to step increases from zero to the first expected occurrence of the next sample, average latency should decrease as less and less time elapses from the input to the next sample. As the time from saccade to step increases through the neighborhood of the next sample, average latency increases as a greater percentage of inputs must wait for the second sample to cause a response. As the time from saccade to step further increases above the expected occurrence of the first sample, latency again decreases as all inputs must wait for the second sample and the time from input to next sample is decreasing as it was for small delays. Thus a nonsynchronized sampled data system predicts decreasing, then increasing, and then decreasing again latency trends as time from saccade to step increases.

Four such saccade-synchronized experiments were performed at the MIT man-vehicle laboratory. Experiment 1 was a returning pulse in which the return of the target was synchronized to the occurrence of a saccade in response to the initial step. Experiment 2 was a continuing pulse in which the second outward step was synchronized to the occurrence of a saccade in response to the initial step (see fig. 10). Experiment 3 was a ramp with inward step in which the occurrence of the inward step was synchronized to the occurrence of a saccade in response to the initial ramp. Experiment 4 was a ramp with outward step in which the occurrence of the outward step was synchronized to the occurrence of a saccade in response to the initial ramp. These experimental inputs were mixed at random and presented to six subjects. For a complete description see Forster (ref. 21).

Figure 6 shows latencies, in response to the synchronized step of the first two experiments, plotted as a function of time after the last saccade. Plotted on the same graph are the predictions of the revised sampled data model with target synchronized sampler control logic ($\tau_{\min}=200$ msec, $T_1=100$ msec, $L_{\text{ave}}=250$ msec) and with nonsynchronized sampler control logic with samples uniformly distributed between 150 and 250 msec. The results of experiments 3 and 4 were inconclusive, showing no meaningful variation of latency as a function of time after a saccade.

The longer latencies of the returning pulse, compared with the continuing pulse, are in agreement with the results of Beeler and are probably due to muscle hysteresis nonlinearities and viscoelastic elements in the eye. The fact that each response is modeled by a different sampler control logic suggests a system with nonsynchronized sampling in which the eye can adjust the sampling slightly in phase and frequency according to observed target movements. This system can model both the increasing portion of latency distributions by shortening the sampling interval and Wheelless's pulse-step results by lengthening the sampling interval.

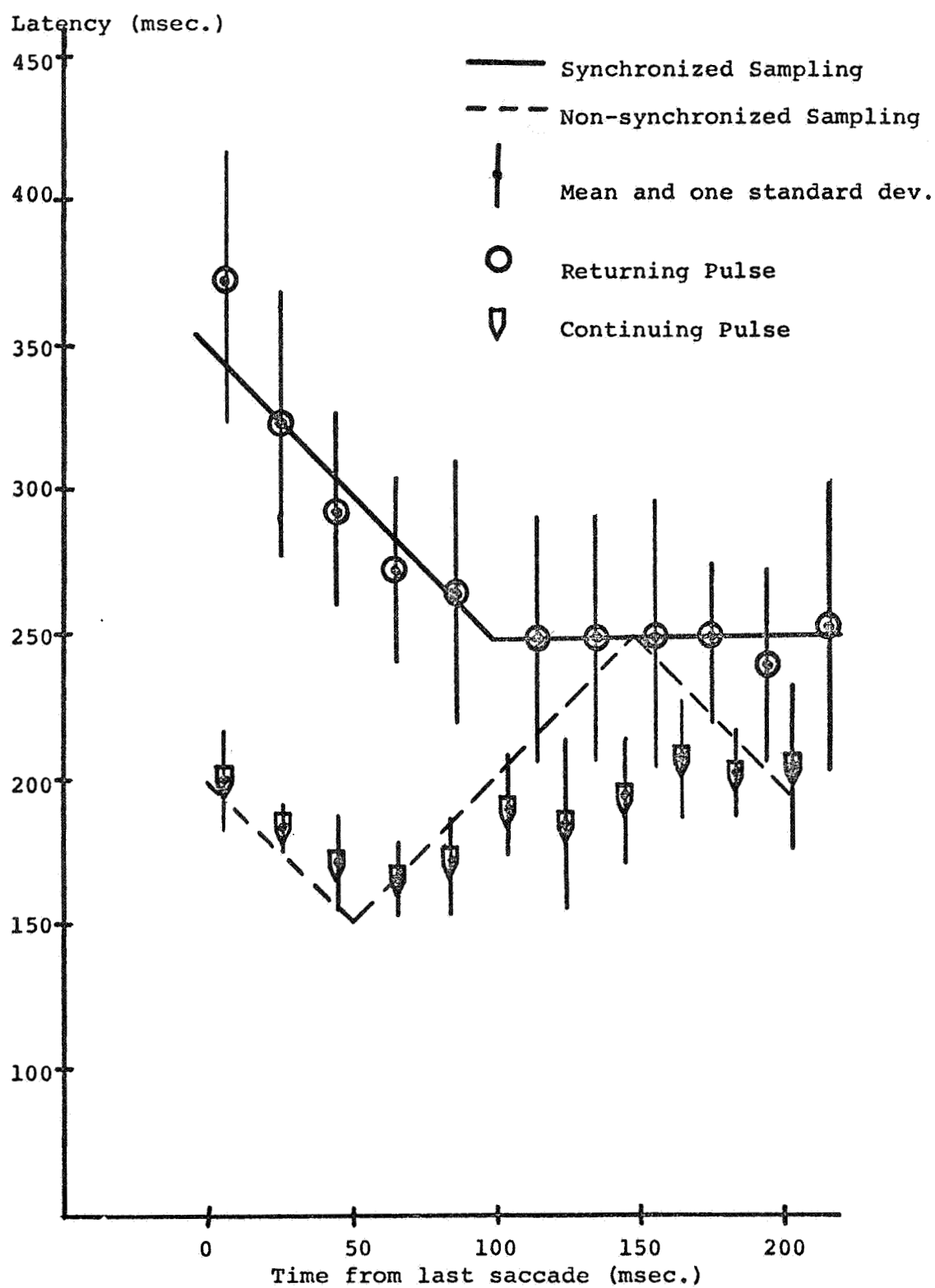


Figure 10.—Results of two saccade-synchronized experiments.

DISCUSSION

Responses of the revised sampled data model have been shown to agree with observed classes of eye movement including steps, pulses, ramps, and step-ramps. In particular, the percent of forms of responses observed in actual eye movements to identical stimuli has been shown to agree with percents predicted by uniformly varying the synchronization of input to sampling instants in the revised sampled data model. The frequency response has been shown to be nearly that of the eye movement control system. The revised model agrees with the findings of Rashbass (ref. 4) that the pursuit eye movement control system is independent of the saccadic control system and proportional to target rate only. The system also agrees with Robinson's findings that the pursuit eye movement control system is continuous. Further, the revised model does not predict pursuit movement in response to small steps as the original model did.

Saccadic inaccuracy and variation in pursuit movement may be modeled by considering the slopes of the nonlinearities stochastically variable. A slope greater than one would produce overshoot, and a slope less than one would produce undershoot. This variability could explain the pursuit velocity exceeding the target velocity in some of the responses observed by Robinson. Further study of these characteristics is necessary to describe the stochastic processes involved.

The revised sampled data model also includes some predictive effects as seen in the agreement of the two sampler control logics with the results of Zuber. The eye shows much more predictive ability than this, however, as seen in the eye's ability to suppress a saccade and increase pursuit latency in various step-ramp experiments. The revised sampled data model predicts this pure pursuit response in a region of step-ramps very close to the region of observed pursuit-only response (ref. 21). This prediction, however, requires that no sample occur in approximately the first 134 msec after the onset of target motion, an event that has a low probability. The model also does not predict the observed longer pursuit latency. Since many no-saccade responses are observed, there is probably a predictive function involved even in nonpredictive tracking.

Although transient responses are well modeled by uniform variation in synchronization of input to sampling instants, synchronization does not explain some other characteristics of eye movement. For example, a nonsynchronized sampled data model cannot predict increasing frequency distributions for step latency. The saccade-synchronized experiments show that it is necessary to change sampler control logic depending upon the input to model the effects of muscle dynamics on latencies to steps immediately following saccades. Further study in saccade-synchronized experiments should include longer times from saccade to input and a step-step-ramp experiment in which the step-ramp is saccade-synchronized.

Impulsive sampling also does not model the pulse-step results of Wheelless, but an extension to finite width sampling is compatible with Wheelless's results. If, for example, a finite width sample occurred during the target step from plus to minus, the sampled error would contain both positive and negative parts that could cancel each other and thus delay the saccadic response until after the next sample. If the pulse width were 40 msec and the average intersample time 200 msec, about 20 percent of samples would cross the plus to minus step (as described above), causing an overall increase in average response time of 40 msec, in good agreement with Wheelless's results.

Subject to the limitations and extensions discussed, the revised sampled data model has been shown to predict responses resembling eye tracking movements with remarkable accuracy. It therefore is a useful tool in predicting eye movement response based on target movement only.

instants, and we divide this interval again in 10. (The parameters used in this sketch are slightly different from the ones mentioned earlier, for simplicity of presentation.) If the sample occurred just before A, there is an initial error AB, which decreases according to BC_i (slope varying with the ramp size). After the pursuit delay time (line XX) the error remains constant (line D_iE_i). At the next sample the error is sensed, and if it exceeds threshold, it is reduced to zero, following a saccadic delay. (Jump from line D_iE_i to reference line ZZ, if D_iE_i is out of zone M.) For a larger input synchronization delay, the error duration becomes shorter (as seen in figs. 5 and 6). If the sample then occurs at a time prior to the pursuit delay, the saccadic jump will overcorrect the error. The size of the jump is read on the scale P, and is applied at a moment read on the scale S. If the correction brings forth an offset in position larger than the dead zone characteristics, a second saccadic correction will appear. (The moment at which a second saccade will show up is read from the sketch at the crossing of the line BD_i and the zone N, the latter being fixed to the line D_iE_i .) For example, if the latency between input and a sample is 0.7, the first saccade will occur at 0.7 on the scale S, and the size of the jump is read at 0.7 on the scale P. Since the size of the saccade is so large that it falls outside the zone N (step of -2 degrees, ramp of 10 deg/sec), a second corrective saccade at the next sample is required to bring the error to zero (compare with fig. 6).

With this figure, one can analyze the chances of having none, one, or two saccades, and one can evaluate their probability, under the assumption of uniformly distributed sampling, for any combination of step-ramp inputs. The diagram has saccadic delay T_1 , pursuit delay T_2 , ramp size, and step size as parameters. For the latter, one shifts zone M and reference line QQ up or down. The zone N is fixed to the line D_iE_i , resulting from the ramp size and the pursuit delay. Being in the zone M means no saccade at all. Being in the zone N means only one saccade, and outside this zone N means two saccades. Their magnitude is read at the entry on the scale P, and applied at the entry on the scale S.

REFERENCES

1. Beeler, G. W.: Stochastic Processes in the Human Eye Movement Control System. Ph.D. thesis, California Institute of Technology, May 1965.
2. Robinson, D. A.: The Mechanics of Human Smooth Pursuit Eye Movement. J. Physiol., vol. 180, 1965, pp. 569-591.
3. Young, L. R.: A Sampled Data Model for Eye Tracking Movements. Dept. of Aeronautics and Astronautics, MIT, June 1962.
4. Rashbass, C.: The Relationship Between Saccadic and Smooth Tracking Eye Movements. J. Physiol., vol. 159, 1961, p. 326.
5. Vossius, G.: Der Kybernetische Aspekt der Willkurbewegung. Progress in Bio-cybernetics (Japan), vol. 2, 1965, pp. 111-140.
6. Zuber, B. L.: Physiological Control of Eye Movement in Humans. Ph.D. thesis, Dept. of Nutrition and Food Science, MIT, June 1965.
7. Wheelless, L. L.: The Effects of Intensity on the Eye Movement Control System. Ph.D. thesis, Univ. of Rochester, 1965.
8. Jury, E. I.; and Pavlidis, T.: Discussion of a Sampled Data Model for Eye Tracking Movements. Second International Congress of the International Federation of Automatic Control (Basel, Switzerland), Butterworths, Ltd., Aug. 1963, p. 461.
9. Johnson, L. E.: Human Eye Tracking of Aperiodic Target Functions. Systems Research Center, #37-B-63-8, Case Institute of Technology.
10. Meiry, J. L.: The Vestibular System and Human Dynamic Space Orientation. Sc.D. thesis, MIT, Dept. of Aeronautics and Astronautics, June 1965.

11. Sugie, N.: The Mechanism of Eye Movements. Bull. Electrotech. Lab. (Japan), vol. 30, no. 7, July 1966.
12. Dallos, P. J.; and Jones, R. W.: Learning Behavior of the Eye Fixation Control System. IEEE Transactions on Automatic Control, vol. AC-8, no. 3, July 1963.
13. Michael, J. A.: An Analysis of a Predictive Component in the Human Visual Tracking System. M.S. thesis, McGill University, Apr. 1964.
14. Lässig, P.: Critical Observations on the Study of the Eye-Following System with Unpredictable Movements of the Aiming Object. (Ger.) Acta. Biol. Med. Ger., vol. 14, 1965, pp. 618-29.
15. Milhorn, H. T.: The Application of Control Theory to Physiological Systems. W. B. Saunders & Co., 1966.
16. Milsum, J. H.: Biological Control Systems Analysis. McGraw-Hill Book Co., Inc., 1966.
17. Jones, G. M.; and Milsum, J. H.: Spatial and Dynamic Aspects of Visual Fixation. IEEE Transactions of Bio-Medical Engineering, vol BME-12, no. 2, Apr. 1965, pp. 54-62.
18. Fuchs, A. F.: Saccadic and Smooth Pursuit Eye Movements in the Monkey. J. Physiol., vol. 91, 1967, pp. 609-631.
19. Saslow, M. G.: Latency for Saccadic Eye Movements. J.O.S.A., vol. 57, no. 8, Aug. 1967, pp. 1029-1033.
20. Westheimer, G.: Eye Movement Response to a Horizontally Moving Visual Stimulus. A.M.A. Arch. Opthal., 1954, pp. 932-943.
21. Forster, J. D.: A Stochastic Revised Sampled Data Model for Eye Tracking Movements. S.M. thesis, MIT, Dept. of Aeronautics and Astronautics, June 1968.

32. Modeling the Human Operator With Finite-State Machines

*Lawrence J. Fogel and Roger A. Moore
Decision Science, Inc.*

The prospect of using finite-state machines to represent arbitrary analog transducers and the human operator in his performance of flight control is considered in this experimental investigation, together with the use of evolutionary programming as a means for finding such representations. A series of tasks of increasing difficulty which demonstrate this capability are presented, ranging from the characterization of relatively simple linear analog "pilots," through more complex nonlinear analog "pilots," and on to the human pilot. Consideration is given to the tradeoff of worth in terms of the adequacy of representation against the resulting complexity of the input and output functions in terms of the size of the alphabet and the complexity of characterization of the logic of the given transducer. The findings of the investigation demonstrate that finite-state machines do offer a suitable means for representing the human operator in terms of his stimulus-response behavior with respect to nonconvergent tracking tasks.

INTRODUCTION

There have been many attempts to obtain suitable models of human behavior. In recent years, these efforts have included attempts to develop mathematical representations with the hope that these might facilitate the design of equipment systems which include the human operator. From the engineering viewpoint, it is of interest to inquire as to the worth of modifying plant dynamics in order to make it more suitable for human control: Will the expected errors of human control seriously degrade system performance? Might these errors introduce undesirable modes of response? and so forth.

At the same time, there is interest in gaining a deeper understanding of the human operator. Having such knowledge should facilitate introducing him into the vehicle system in a most meaningful way. In this regard it is important to resolve such questions as: How does the human operator organize his perception of the dynamic plant under his control? How well does he know the criteria against which he optimizes his performance? To what extent are these criteria affected by the displays and controls, to say nothing of the instructions he receives? What kind of displayed information would be most suitable in terms of the desired performance of the system? and so forth. Traditionally, the search for adequate models of the human operator began with linear dynamic representations which were generally

expressed in the form of differential equations; this, in spite of the fact that it was universally recognized that the human operator is nonlinear in his transduction. Hopefully, a best linear representation would account for a major portion of the observed behavior and leave only a small residue of unaccountable deviations. Although such linear transfer representations of the human operator have proven suitable for most engineering purposes, specific applications often require considerable reduction in the amount of residual error which usually remains. Unfortunately, the amount of this remnant generally depends upon the dynamics of the plant and upon the particular human operator under consideration. Doubt will remain as to the legitimacy of interpreting the model in any psychological sense until it becomes possible to distinguish that component of the remnant which is strictly random from the remainder which is due to the particular nonlinearity of the human operator.

A key to this problem may lie in the use of a different class of mathematical functions for modeling the human operator. A convenient mathematical function for this purpose is the finite-state machine—a machine being defined in terms of the finite alphabet of possible input symbols, the finite alphabet of possible output symbols, and some finite number of internal states which the machine may assume. To specify such a machine, each of these states must be described in terms of the symbols which would emerge from the machine and the next state the machine would assume when the machine is in any state and receives each of the possible input symbols. For example, a three-state machine is shown in figure 1. Here the alphabet of input symbols is comprised of 0 and 1, whereas the alphabet of output symbols is comprised of α , β , and γ (input symbols being conventionally placed to the left of the virgule while output symbols are placed to the right). As shown in the state diagram, an input of the symbol 1 while the machine is in state B will cause the machine to output the symbol α and change its internal state from B into C. On the other hand, an input of the symbol 0 when the machine is in state B causes an output symbol γ and the machine remains in state B. Thus a sequence of input symbols is transformed into a sequence of output symbols, as shown below, assuming that the machine is in state C when the first input symbol is received. In general, the output sequence would have been different had the machine been in another initial state. Thus it is seen that a finite-state machine is completely specified by the state diagram and an identification of its initial state. It is assumed that the machine acts only when an input symbol is received and that the operation is always completed before the next input symbol is received.

Present state	C	B	C	A	A	B
Input symbol	0	1	1	1	0	1
Next state	B	C	A	A	B	C
Output symbol	β	α	γ	β	β	α

In essence, a finite-state machine representation of the human operator portrays his behavior as being dependent upon his internal (psychological) state. For each such state, he is presumed to behave in a consistent manner with respect to the stimulus-response and with respect to the transition to each next state on the basis of the received stimulus. For example, if the pilot is in the indifferent state for some class of stimuli, his response is nil and he remains in that state. For some other class of stimuli, his response may also be nil, but any one of these will transfer him into a new state—a responsive state. Still another class of stimuli will place him in a responsive state and, at the same time, will elicit a particular output symbol (his overt reaction).

Clearly, any finite organism operating upon a finite set of discriminably different stimuli and responding within a finite set of discriminably different output symbols can be exactly represented by a finite-state machine. The hope is that actual human performance can be represented in terms of a relatively simple finite-state machine, thus depicting the

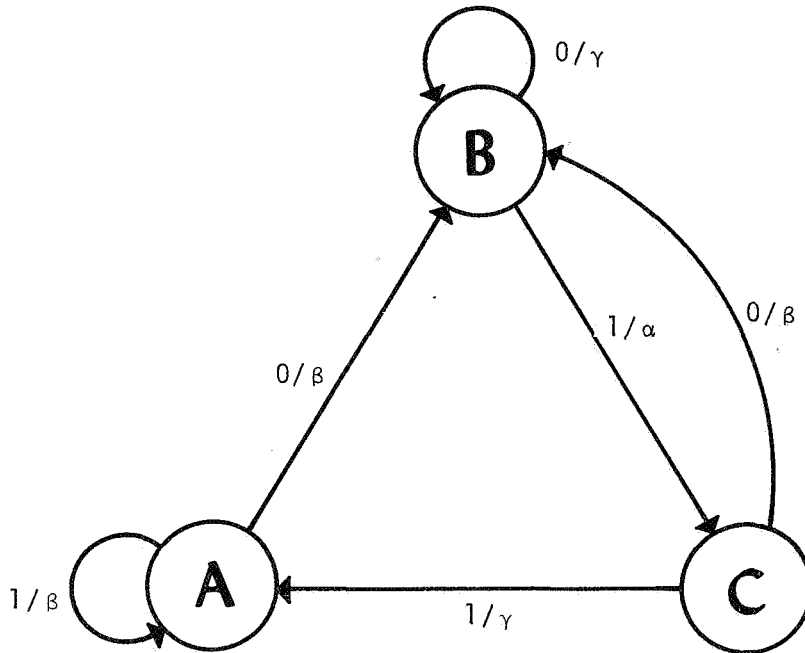


Figure 1.—State diagram for a finite-state machine.

essential ingredients of human behavior, and this in an intuitively satisfying manner. That is to say, given a finite sequence of input-output data, it is always possible to find a finite-state machine which will exactly represent the given transduction. The problem is to find a meaningful approximation to such a machine in the sense that this replication reflects not only the measured input-output behavior of the human operator but also his goal-seeking intent. Such a model might be expected to reveal important consistencies of human behavior, not only as these are evidenced in terms of overt transduction, but also in terms of state properties having psychological significance.

A simplified evolutionary program was used for determining such models on the basis of the empirical evidence—the sequence of input and output symbols; that is, the observed stimuli and response (ref. 1). More specifically, in this program the search was restricted to a special class of finite-state machines structured such that an ordering can be given to the states in the following manner: For any state, consider inputs x and y with $x > y$. If under x the transition is to state A and under y the transition is to state B , then $A \geq B$. In essence, this consists of using some prior knowledge regarding the physical domain under investigation. However, the search is still of a highly general nature even though it is made more efficient by the constraint of possible structure of the finite-state machines.

DISCUSSION

The above-described technique was used to model a linear dynamic system, a nonlinear dynamic system, and a human operator in single-axis tracking. In addition, empirical evidence derived from actual flight performance of a pilot was examined. Some of the results of these experiments are described on the following page.

A linear dynamic system of the form

$$\frac{K_1 A + K_1 K_2 s}{(A + s)^2}$$

is commonly used to model the human operator in tracking, the constants being chosen to typify the human operator in control of a second-order plant having the transfer characteristic K/s^2 through the use of a compensatory display. The input and output of this system were recorded as the system was driven by a random disturbance, these data being quantized into 64 equal class intervals over the amplitude range.

Modeling of this analog pilot through the use of finite-state machines operating upon discrete input data can be compared to numerical integration wherein each sensed amplitude is taken to correspond with a rectangular segment of that specific height. If the time interval between samples is too great, there will be an error due to the poor approximation of the step function at each point in time. On the other hand, if the time interval becomes too short, the round-off error becomes predominant. In this case, the range over which the time interval between data points can be varied was rather severely curtailed in view of the restriction to a 64-symbol alphabet. That is to say, amplitude quantization places an important constraint upon the limit of the longest meaningful time interval of sampling. A sampling time interval of 0.12 second was considered to be most appropriate, that is, introduction of insignificant error.

The evolutionary program may be expected to produce different results for different definitions of best. Table 1 indicates the result of three experiments performed on the same data base of 4 000 points (these being individually generated from the analog record every 0.12 second apart). In each case, a best finite-state machine was found and driven with a sequence of 5 000 input symbols, consisting of the above-referenced 4 000 data points and an additional 1 000 data points, in order to test the adequacy of the model in fitting new data.

In the first experiment, a magnitude-of-the difference scoring criterion was used; that is, the error for each data point was taken to equal the number of symbols between the output of the finite-state machine and the actual data point. In the second experiment, an all-or-none scoring was used; that is to say, the error was taken to be unity if the output of the finite-state machine differs from the actual output, and 0 if it does not. Note that with this more stringent criterion, the number of correct estimations increases from 33.1 percent to 36.26 percent; on the other hand, as might be expected, larger errors occur more frequently. In the third experiment, the output symbols of the finite-state machine were set by

TABLE 1

Exp. No.	Criterion	Average Error per Data Point	Cumulative Error					
			0	<1	<2	<3	<4	<5
1	Error Magnitude	1.35	33.10	61.12	81.00	92.58	97.90	99.52
2	All-or-None	1.53	36.26	57.40	75.06	86.76	94.36	98.14
3	rms Error	1.38	24.74	60.16	83.76	94.96	98.90	99.82

averaging. This is equivalent to accepting a least square's criterion. Under this criterion, the correct estimate score was reduced to 24.74 percent; and there was a significant reduction in the number of large errors. Figure 2 indicates a sample of the results of this third experiment.

Table 2 indicates the relationship between the size of the finite-state machine and the average error in fitting these same data points. As expected, the greater the number of states, the less the average error.

TABLE 2

<u>Exp. No.</u>	<u>No. of States</u>	<u>No. of Exercised State-Input Pairs</u>	<u>Average Error per Data Point</u>
4	7	258	1.76
5	15	473	1.46
6	31	909	1.35

It is also of interest to consider the effect of alphabet size, the adjusted accuracy being defined as the alphabet resolution multiplied by the average error measured in terms of the number of symbols. The expected accuracy symbol for a 64-symbol alphabet, given a uniform distribution, is 1.56 percent. This figure, multiplied by the average error associated with a given machine, yields the adjusted accuracy. Figure 3 indicates the effect of machine size upon adjusted accuracy for an input alphabet of 64 symbols. Note that little improvement can be expected in going beyond 31 states.

A second series of experiments was devoted to the modeling of a nonlinear dynamic transducer representing the pilot, its logic being: When the motion of the plant is away from the reference, the pilot responds in proportion to the rate of this motion. If the plant is moving toward the reference, there is no response by the pilot unless the rate is such that a simple linear prediction indicates that the plant would cross the reference in less than T seconds. In this case, the pilot responds in proportion to the rate of motion. All outputs of the system are transformed by a second-order approximation to a transport delay analog filter to represent the inertia of the pilot's musculature. The input and output of this nonlinear transducer were recorded as it caused a second-order plant to track a random variable.

Cursory examination of the best-fit finite-state machine indicated that its output led the actual output by a noticeable amount. It was therefore, considered worthwhile to introduce a delay in the output, this corresponding to the reaction time of the human operator. This was done by fitting the $(n+d)$ th output symbol with the input symbol n , where d is the delay in an integral number of sampling intervals. Table 3 summarizes these experiments. While the scores for the 31-state machine do not show a dramatic change with the introduction of delay, the scores for the 1-state machine do change drastically.

Further experiments were performed using the 0.30 second delay in that the best scores for the 1-state machine and the 31-state machine were for this same delay. Table 4 indicates the results of these experiments.

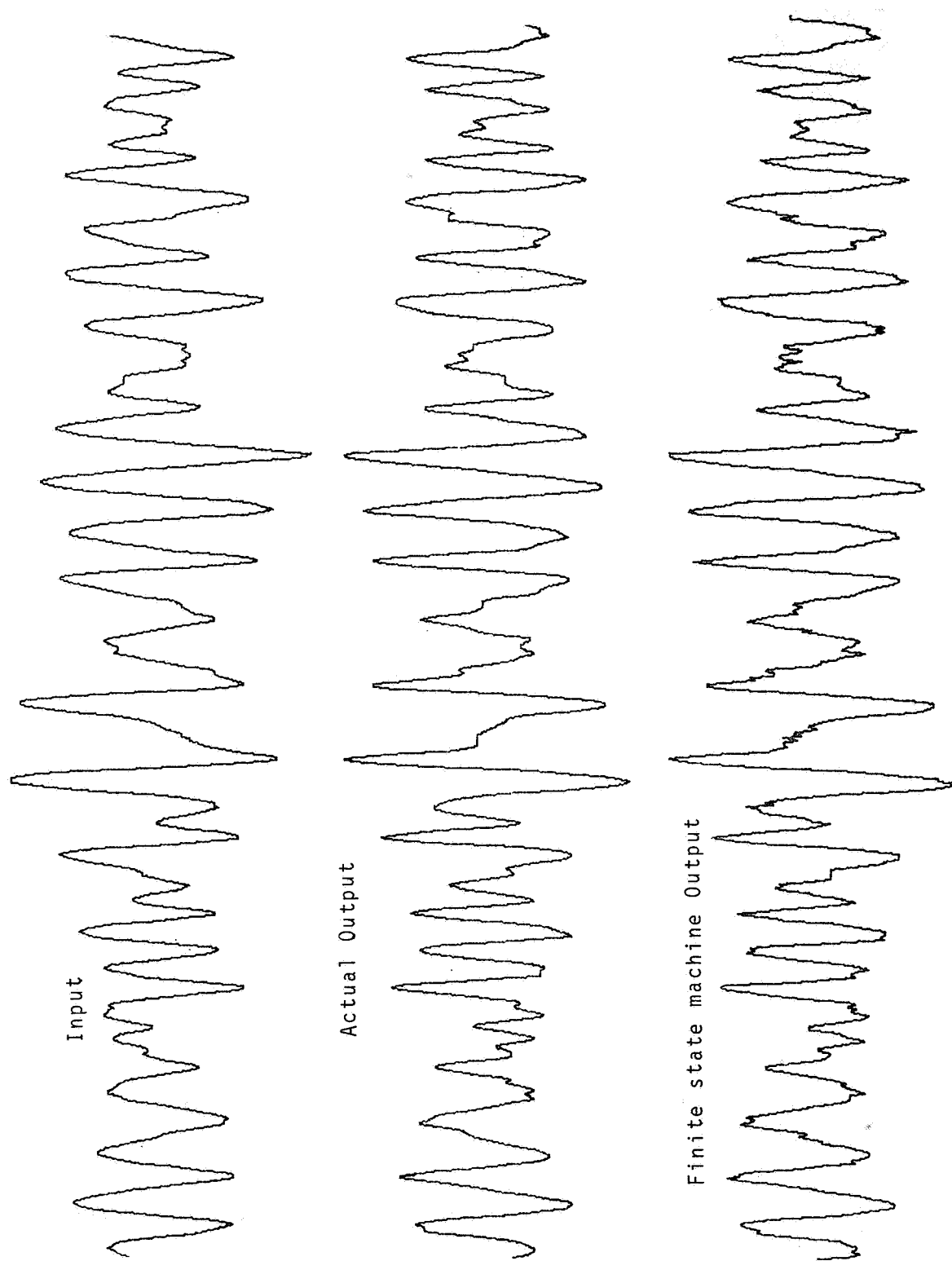


Figure 2. — Comparison of linear dynamic system with 31-state machine (experiment 3).

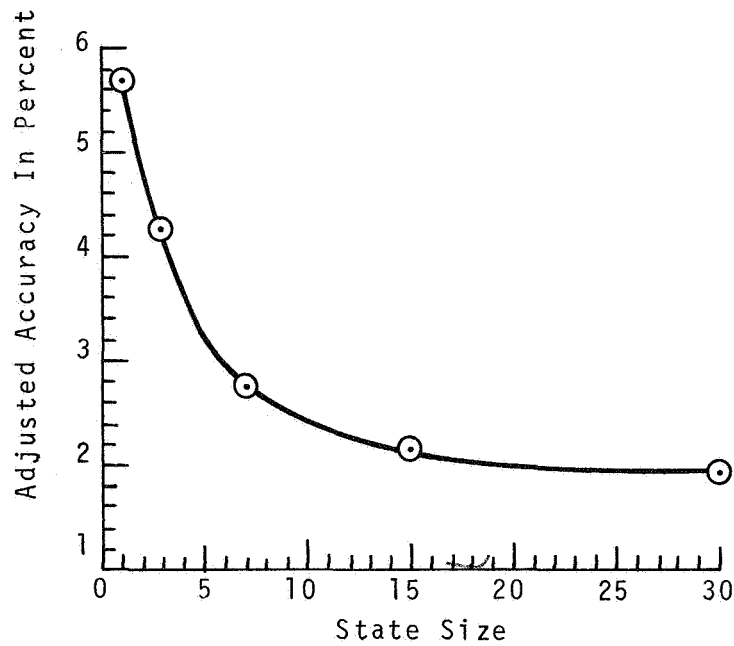


Figure 3.—Adjusted accuracy versus state size for input alphabet of 64 symbols.

TABLE 3

Exp. No.	Delay in Seconds	Average Error for a 31-State Machine	Average Error for a 1-State Machine
7	0.00	1.72	6.34
8	0.15	1.73	4.48
9	0.30	1.69	2.31
10	0.45	2.13	3.84

TABLE 4

Exp. No.	Criterion	No. of States	Average Error per Data Point	Adjusted Accuracy in Percent
11	rms Error	7	2.01	3.12
12	rms Error	15	1.35	2.11
13	rms Error	31	1.06	1.65
14	Error Magnitude	31	1.01	1.57
15	All-or-None	31	1.12	1.75

With the rms criterion, 37 percent of the symbols were correct as compared with 47.8 percent for the error magnitude criterion and 50.6 percent for the all-or-none criterion. However, 90 percent of the data points were in error by less than 2.5 symbols when the rms criterion was used, whereas 90 percent of the data points fall within 2.9 symbols for the error magnitude criterion; and 90 percent of the symbols fall within 3.5 symbols for the all-or-none scoring. Figure 4 offers a comparison of a sample of the actual output of the nonlinear dynamic transducer, the corresponding output of the finite-state machine model of that transducer, and their difference.

While the behavior of the nonlinear analog pilot was being recorded, the stimulus and response of a human operator was also recorded as he controlled a second-order plant. In this experiment, the human operator tried to adopt the strategy of the nonlinear transducer. The results of various attempts to model the human operator are summarized in table 5. Note that a delay of 0.15-second results in a significant reduction in the error. A sample of the result of experiment 21 is shown in figure 5.

TABLE 5

<u>Exp. No.</u>	<u>No. of States</u>	<u>Criterion</u>	<u>Delay in Seconds</u>	<u>Average Error per Data Point</u>	<u>Adjusted Accuracy in Percent</u>
16	15	Error Magnitude	0.0	3.5	5.5
17	31	Error Magnitude	0.0	3.3	5.1
18	31	Error Magnitude	0.15	2.6	4.0
19	31	Error Magnitude	0.30	3.5	5.4
20	31	rms Error	0.30	3.7	5.7
21	31	rms Error	0.15	2.7	4.2

As expected, the accuracy of modeling the actual pilot was considerably less than that obtained when modeling the analog pilot, that is, the nonlinear dynamic transducer. The human operator was, however, seen to behave in quite a consistent manner as evidenced by the result of experiment 21. The error magnitude criterion produced correct results 28.3 percent of the time while the rms error criterion achieved a comparable score of only 17 percent. As expected, however, the larger errors were significantly reduced through the rms scoring.

Further experiments were performed in attempts to model the human operator with respect to four criteria: error magnitude, all-or-none, rms error, and zero band error magnitude, the last of these being similar to the error magnitude criterion except that no penalty is associated with errors which fall within one symbol of the correct value. These experiments used 31-state machines expressed in a 64-symbol alphabet, there being 0.15-second sampling of the analog record and 0.15-second delay introduced to compensate for human reaction time. In each of these experiments, the structure of the initial finite-state machine was the same as that used in the previous experiments devoted to modeling the analog pilot (table 6).

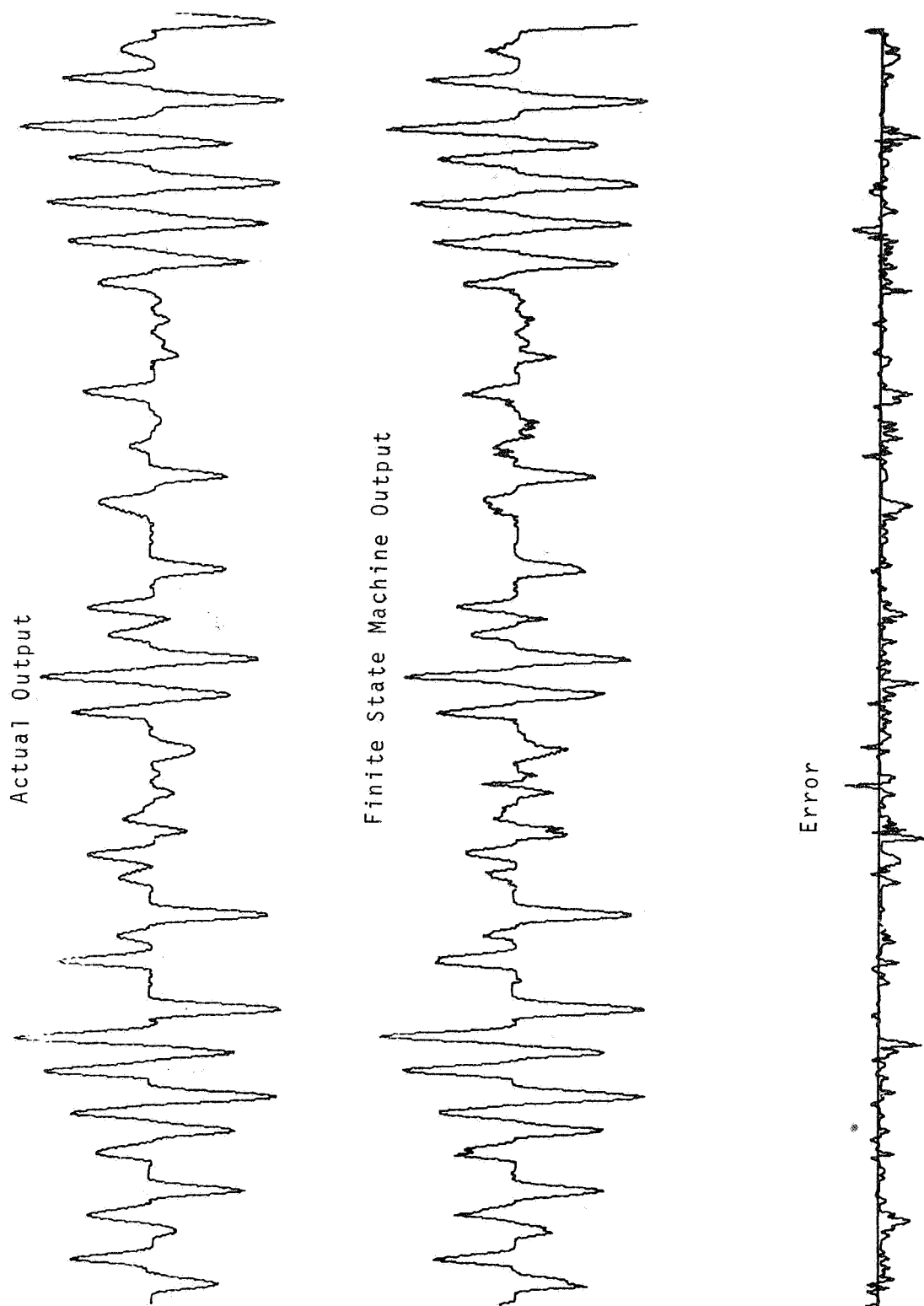


Figure 4. — Comparison of nonlinear dynamic system with 31-state machine using all-or-none scoring (experiment 15).

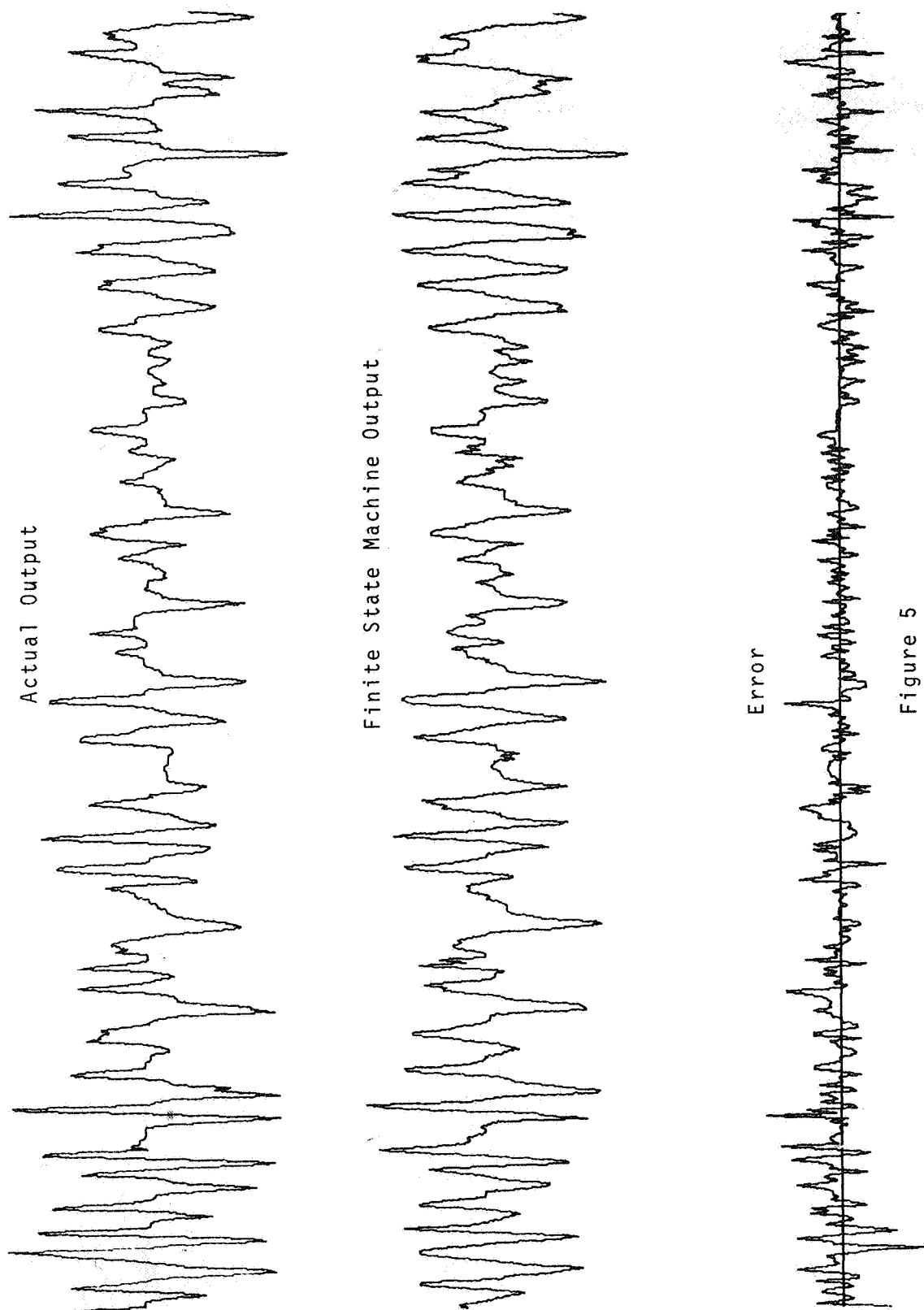


Figure 5

Figure 5. — Comparison of human operator with final 31-state machine (experiment 21).

TABLE 6

<u>Exp. No.</u>	<u>Criterion</u>	<u>Average Error per Data Point</u>	<u>Average Error in Percent</u>
18	Error Magnitude	2.55	3.96
21	rms Error	2.67	4.16
22	All-or-none	3.10	4.83
23	Zero Band Error Magnitude	2.58	4.02

The effect of imposing distinctly different expressions of goal on the distribution of error is shown in table 7. Each criterion is expressed in terms of a different error-cost matrix so that the program determines different output symbols for each state-input pair. The goal in experiment number 18 was to minimize the average error; and, as shown, this experiment produced the smallest average error. The goal in experiment 21 was to minimize large errors, and this was clearly the result. The intent in experiment 22 was to achieve the greatest number of correct predictions under the philosophy of "A miss is as good as a mile." In experiment 23 all errors not in excess of one symbol from being correct were accepted as being correct, the goal being to minimize the average error with respect to all other deviations.

Examination of the results of these experiments revealed the strategy of the pilot to be comparable to that exhibited by the nonlinear analog pilot (as described above). In these experiments there was the concern that the recall used for fitting the machine might be so small that the resulting small remnant may be due to the uniqueness of exercising state-input pairs. Experiment number 24 was conducted to clarify this point. In this experiment,

TABLE 7

<u>Exp. No.</u>	<u>Criterion</u>	<u>Percent Correct</u>	<u>Percent Within 1 Symbol</u>	<u>90 Percentile Point</u>	<u>Average Error in Number of Symbols</u>
18	Error Magnitude	28.3	48.2	7	2.55
21	rms Error	16.9	41.2	6	2.67
22	All-or-none	32.9	47.6	9	3.10
23	Zero Band Error Magnitude	25.0	51.0	7	2.58

the final machine found in experiment 13 was driven by the input used by the human pilot and the output of this machine was compared with the pilot's output. The result was surprisingly good, the average error being 3.65 symbols. While this is not as low a score as that obtained from directly modeling the human operator, it did indeed confirm that the length of the recall was adequate and that the primary strategy of the human pilot was indeed rate control.

Further studies were conducted in order to establish the difference between the finite-state machines obtained by modeling the human operator and those obtained by modeling the nonlinear analog pilot. The result of the first of these experiments is shown in figures 6 and 7. These data were obtained from the final machines found in experiments 13 and 18 by holding the rate constant and plotting the output symbol as a function of the input symbol. For example, the curve for zero rate is a plot of the output symbols for an input of 1 into state 1, an input of 2 into state 2, and so forth. The family of curves described in these figures show that the human operator was not using a pure rate mode of control. This is most clearly shown in figure 6. The nonlinear analog pilot maintained an almost constant output of symbol 32 regardless of the level of input symbol while the human operator maintained a tendency to correct for positional errors. This may be seen in that for all levels less than the midpoint (symbol 32), the human operator's output symbols were above the line derived from the nonlinear pilot model; and for all levels greater than the symbol 32, the human operator's output symbols were below that of the nonlinear analog pilot. In general, all of the remaining curves in this family demonstrate this same characteristic. With larger excursions from the center position, the discrepancy between the model of the human operator and that of the nonlinear analog pilot becomes greater. It is also evident from these data that with larger rates the human operator's output becomes more erratic, as shown in figure 7. In summary, the results of these experiments indicate that the pilot was trying to fly by rate, but that he was also influenced by position. As expected, he demonstrated greater deviations from his strategy at higher rates.

CONCLUSION

The above-described experiments demonstrate that finite-state machines do indeed offer a suitable means for representing the human operator with respect to his stimulus-response behavior in single-axis nonconvergent tracking tasks. In fact, the logical representation expressed in this form serves to produce smaller residual error (remnant) than is achieved through use of more conventional techniques operating on a similar data base. Further, the specific finite-state machine models are derived through a goal-seeking technique which replicates the heuristic behavior of the human operator in the sense that he attempts to minimize some loss function. Thus, additional experiments performed on the same data base, but with respect to different expressions of goal, can be used to indicate an approximate representation for the goal exhibited by the human operator. In point of fact, the strategy of the human operator was clearly discernible on the basis of the finite-state machine modeling of his input-output time history in the above-described experiments.

This investigation marks the initiation of a point of view with respect to representing the human operator by finite-state machines. Clearly, additional work is in order. It is important to examine the dependency of such representation as the human operator reflects different transfer properties of the system under his control. There is also the question of the degree of universality of such representation; that is to say, presumably given a sufficient time history of human performance, it should be possible to achieve a sufficiently precise representation of the individual to uniquely characterize this person. On the other

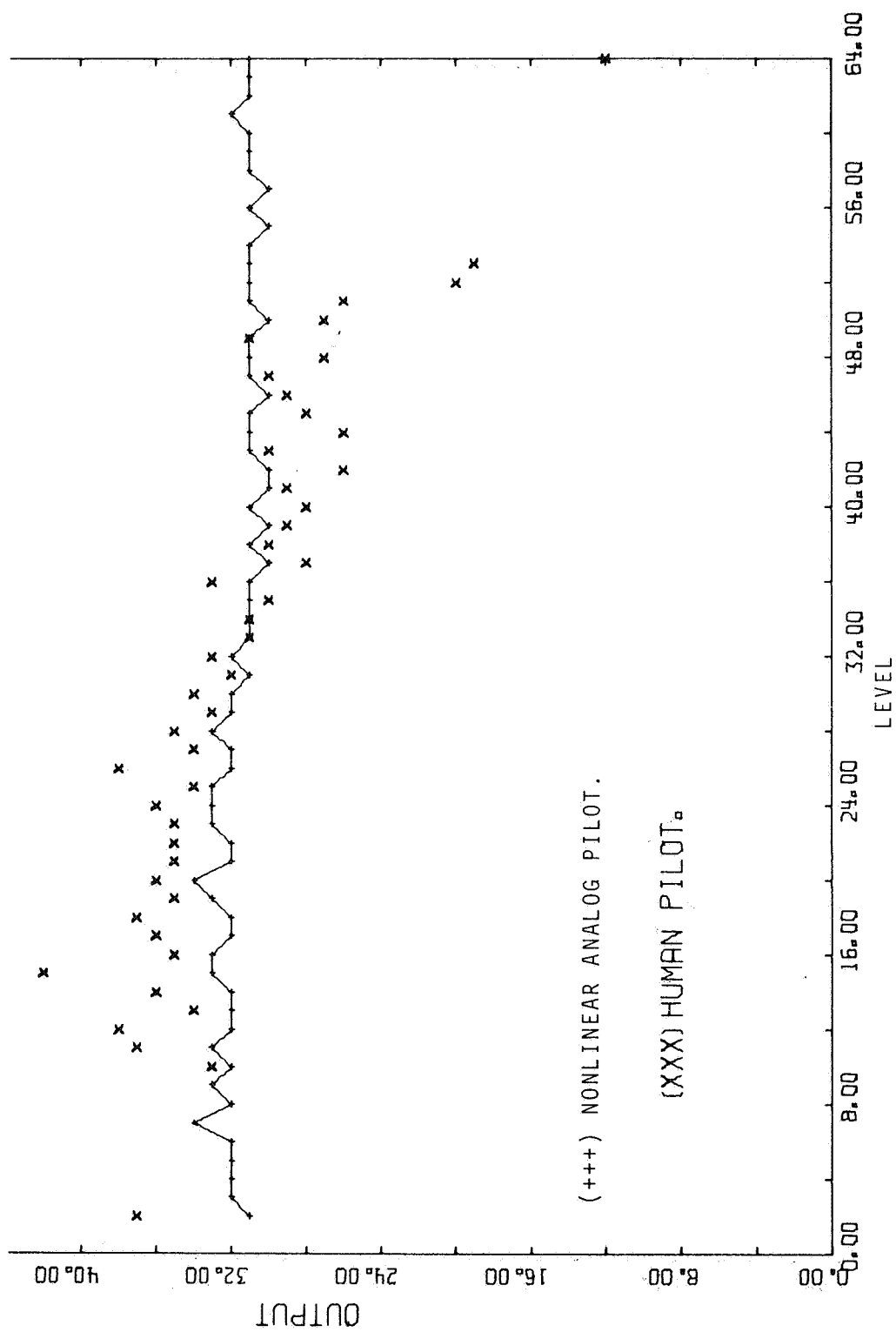


Figure 6. — Comparison of models of human operator and nonlinear analog pilot. Output is shown as a function of level for constant rate of zero.

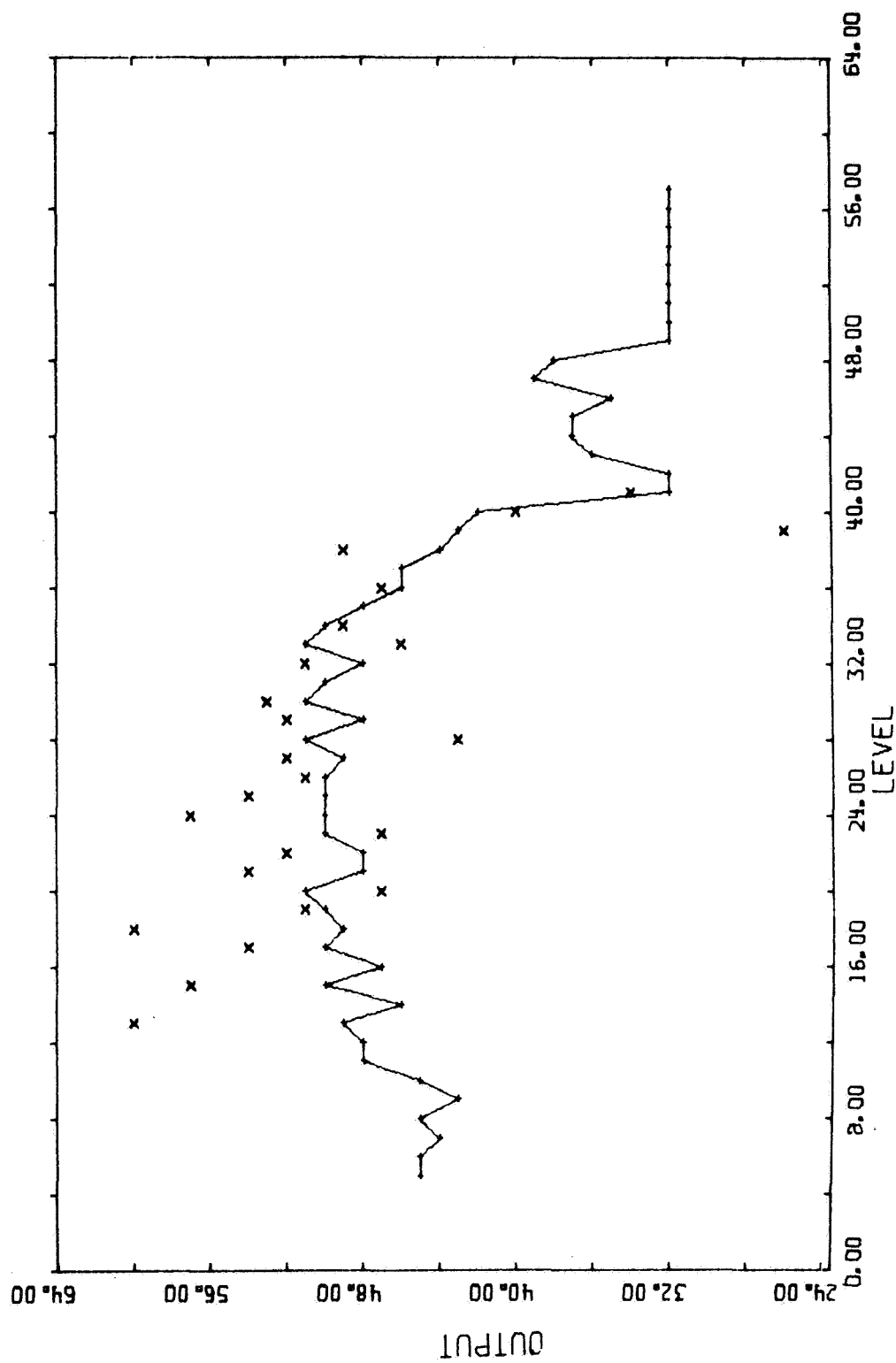


Figure 7.—Comparison of models of human operator and nonlinear analog pilot. Output is shown as a function of level for constant rate of -7.

hand, it should also be possible to find finite-state machines of lesser complexity which are in greater error with respect to the individual but offer a more appropriate means for modeling the class of operators to which this individual belongs. Then, too, it appears appropriate to consider modeling the human operator in convergent tracking tasks (such as occurs in landing of an aircraft). There is also a need to recognize that the human operates in a multivariate situation, thus requiring a higher level of finite-state machine representation — one which includes his switching function among variables.

Motivation for further studies may also be found in the possibility of objectively determining approximation to the human operator's goal on the basis of his exhibited behavior (this goal being distinct from the goal he is instructed to follow, and possibly even the goal he believes he is following). Having such a capability opens the door to more effective evaluation of alternative displays and controls, to assessing the degree of capability of trainees, to designing optimal training sequences and instruction, and to the design of improved control systems for complex man-machine systems.

REFERENCE

1. Fogel, Lawrence J.; Owens, A. J.; and Walsh, M. J.: Artificial Intelligence Through Simulated Evolution. John Wiley & Sons, Inc., 1966.

IX. NEUROMUSCULAR SYSTEM ANALYSIS

33. A Closed-Loop Neuromuscular System Explanation of Force Disturbance Regulation and Tremor Data*

*R. E. Magdaleno and D. T. McRuer
Systems Technology, Inc.*

An adaptive model has recently been evolved to describe small perturbation operations of neuromuscular actuation systems involved in tracking tasks. This model is compatible with both high quality physiological component data and human operator describing function data of low variability and large dynamic range. The purpose of this paper is to use the model to explain force disturbance regulation data and limb tremor data as closed-loop phenomena. This also serves to illustrate a key feature of the model — that the muscle spindles feed back an internal muscle length.

The force disturbance data available in the literature was taken for a large inertia restraint and various values of muscle tone. The response to a torque impulse has a dominant second-order component which is incompatible with an open-loop muscle-load response since this system is third-order. However, the spindle feedback of an internal muscle length results in a numerator zero in the limb response to force disturbance describing function. This zero varies with set tension in a fashion that is nearly the same as the real pole of the muscle-manipulator system. This approximate cancellation thus yields a dominant second-order system with parameter variations with tension which give a qualitative match to the data.

The limb tremor frequency data available in the literature was for variations in spring rate and muscle tone (no added inertia) and for variations in inertia and muscle tone (with no spring). The spindle feedback of an internal muscle length produces a constraint on the lightly damped high frequency dominant mode. The variations of this mode with muscle tone and manipulator restraint values are consistent with the limb tremor frequency variation data.

*This paper includes research efforts supported by the Ames Research Center, Man-Machine Integration Branch, NASA, Moffett Field, Calif. under Contract NAS2-2824.

INTRODUCTION

Recent studies of the neuromuscular system's individual components, as well as the system functioning of ensembles of these components, has led to a quasi-linear model that is compatible with an enormous amount of physiological data as well as human operator input/output describing function data (refs. 1 and 2).

The neuromuscular system model uses control engineering descriptions of the individual components as well as the system functioning of ensembles of these components. The component data of interest include recent anatomical and physiological data for the muscle spindle and input/output studies of the muscle. These data indicate that quasi-linear models can describe the basic behavior of these two elements for small perturbations about an operating point. This model contains three key developments: the variation in muscle system parameters as a function of average muscle tension or operating point, the role of the muscle spindle both as an adaptive equalization element and in its effect on setting muscle tone or average tension, and the spindle feedback of an internal muscle length which constrains some dominant closed-loop roots.

The muscle spindle system also serves to route command inputs from the central nervous system to the neuromuscular system. These inputs are well suited for tracking tasks. However, in this paper we are interested in situations which do not require tracking, but do involve a variety of operating point tensions and manipulator restraints. The data for these situations (torque disturbance regulation and limb tremor) can be explained with the aid of another key spindle feature — that it feeds back an internal muscle length rather than the actual muscle length.

HUMAN OPERATOR DATA INDICATING NEUROMUSCULAR SYSTEM EFFECTS

In this section we are interested in torque disturbance and limb tremor data for a variety of operating point tensions and manipulator restraints.

TORQUE DISTURBANCE REGULATION.—The effect of average voluntary muscle tension on the wrist twisting (supination/pronation) response to torque disturbance inputs has been investigated in references 3 to 6. (The ref. 3 work is reviewed in ref. 7.)

In reference 3 the relative amount of muscle tension was inferred from a sphygmomanometer cuff attached around the forearm. This was displayed to the subject who then could readily set the reading to any one of five levels. These experiments were carried out using irregularly spaced mechanical impulses delivered without warning by a pendulum. The manipulator restraint consisted of an inertia several times larger than that of the arm, and the subject was asked to resist the perturbing influence of the pendulum-produced disturbance on the load.

The transient response resembled that of a dominant second-order system with light damping (fig. 1(a)). Figure 1(b) shows similar results from reference 5 where the disturbance torque was delivered by a motor attached to the shaft of the handle which the subject grasped. A transducer sensed the grip force which is assumed to be indicative of muscle tone in the participating muscles.

Figure 2, adapted from reference 3, shows the general location of the upper pole position of a damped second order fitted to the transient response for five tension values. In general, increasing mean tension increases the natural frequency of these roots but leaves

the damping ratio relatively unchanged. This trend is summarized in figure 3 in terms of the effective time constant¹ of these roots as a function of inferred tension.

EFFECTS OF MUSCLE TENSION AND MANIPULATOR RESTRAINTS ON LIMB TREMOR.—Limb tremor has been observed in a variety of situations. It typically appears in the time history as an oscillation or in the power spectral density as a significant peak at high frequency. Tremor frequencies are often in the range of 6 to 10 hertz as in the analysis

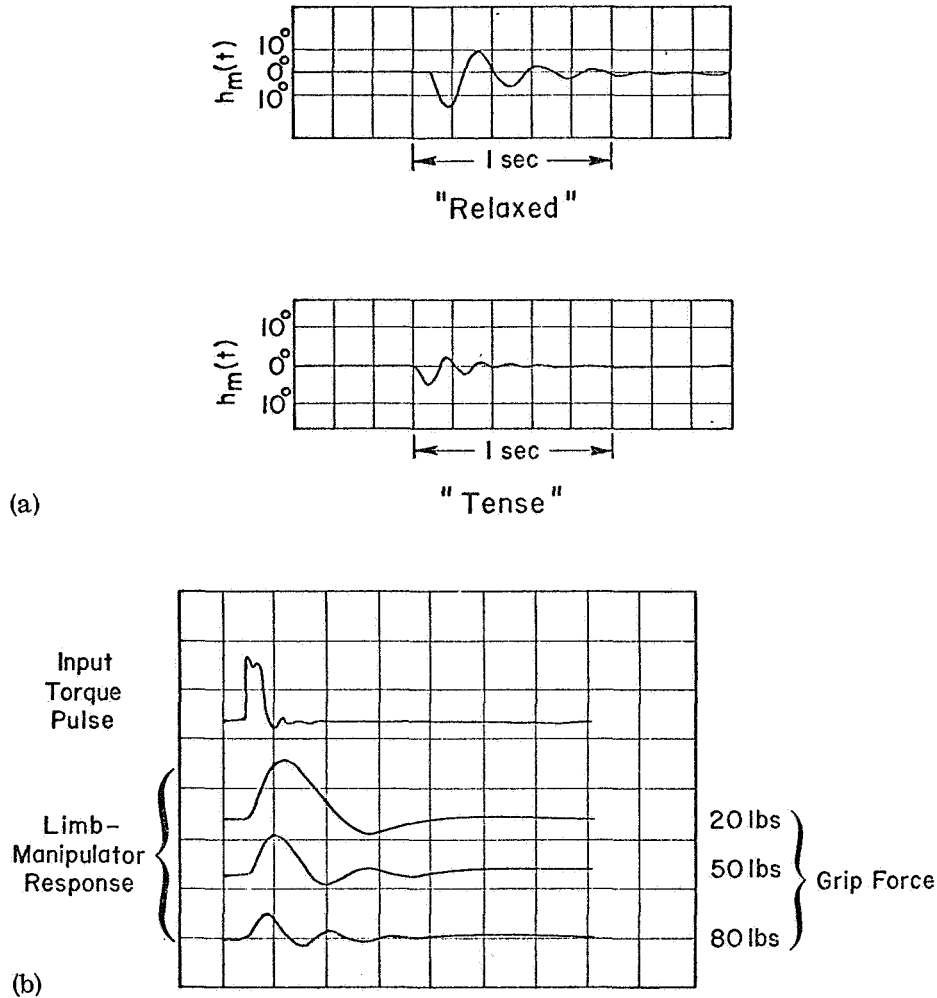


Figure 1.—Responses to a torque disturbance for various tension levels. (a) Response to pendulum input (adapted from ref. 3). (b) Response to a torque motor input (adapted from ref. 5).

¹The effective time constant for a second-order system is given by

$$\frac{1}{1 + \left(\frac{2\zeta_N}{\omega_N}\right)j\omega + \left(\frac{j\omega}{\omega_N}\right)^2} \doteq e^{-(2\zeta_N/\omega_N)j\omega} ; \quad \omega < \omega_N$$

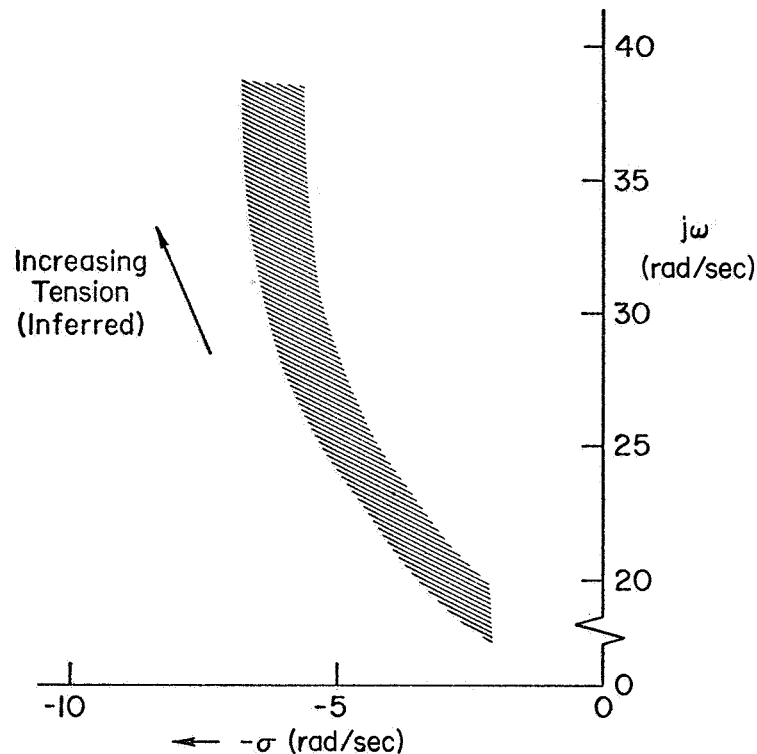


Figure 2.—Upper root location of a complex pair fitted to some transient responses (ref. 3).

of finger velocity (ref. 8) and with different load masses (ref. 9). Generally, muscle tremor frequencies depend on both the manipulator restraint and muscle tension (refs. 10, 11, and 12). In figure 4, forearm flexor tremor frequency is plotted as a function of the mean tension (ref. 10). This tension was exerted against a spring (attached to the wrist) which was oriented parallel to the upper arm. For each spring, increasing the tension causes a slight increase in tremor frequency except at the higher tensions where the data level off. However, an increase in spring rate produces an incremental increase in tremor frequency which is essentially independent of tension. Reference 12 had results similar to those in figure 4, except that only one spring rate was used.

In reference 11 the task was to track a ramp input (pursuit display). The subject gripped a handle in his fist and using wrist rotation about the forearm axis was able to generate his response. Four different load inertias were used. In addition, muscle tension was inferred from handle grip tension as sensed from a rubber pneumatic balloon and pressure-voltage transducer. The dependence of tremor frequency on inertia and inferred muscle tone is given in figure 5. For each inertia there is an increase in frequency as tone increases, whereas for constant tone an increase in inertia causes a decrease in frequency.

Additional results in reference 11 indicate that the tremor frequency of one hand is independent of the other. Specifically, if one hand is controlling a large inertia without pressure, then its tremor frequency can be much less than that of the other hand if it is

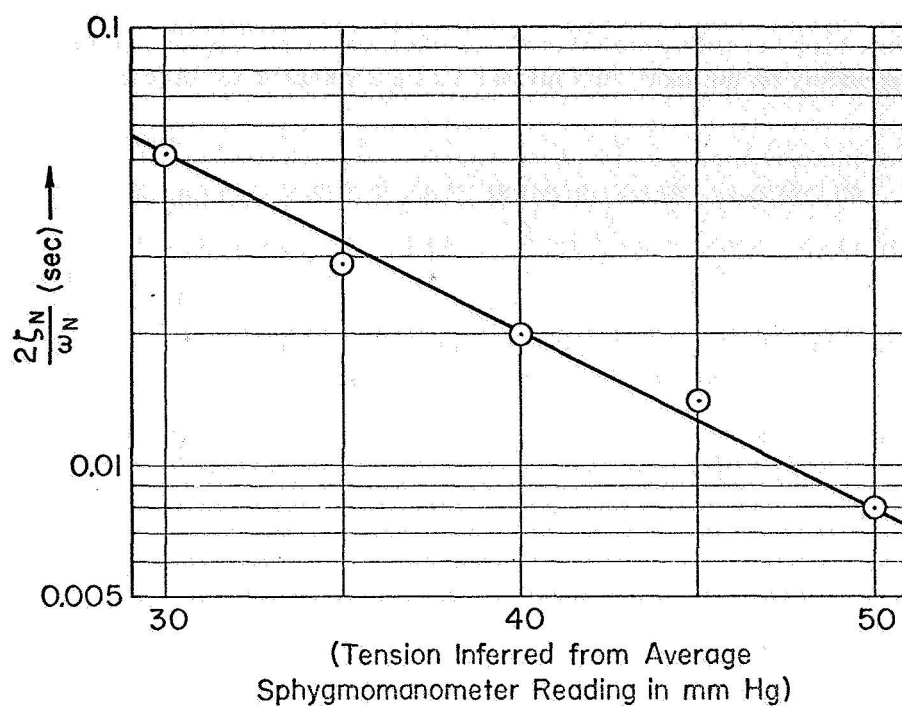


Figure 3. —Effective time constant as a function of inferred tension (ref. 3).

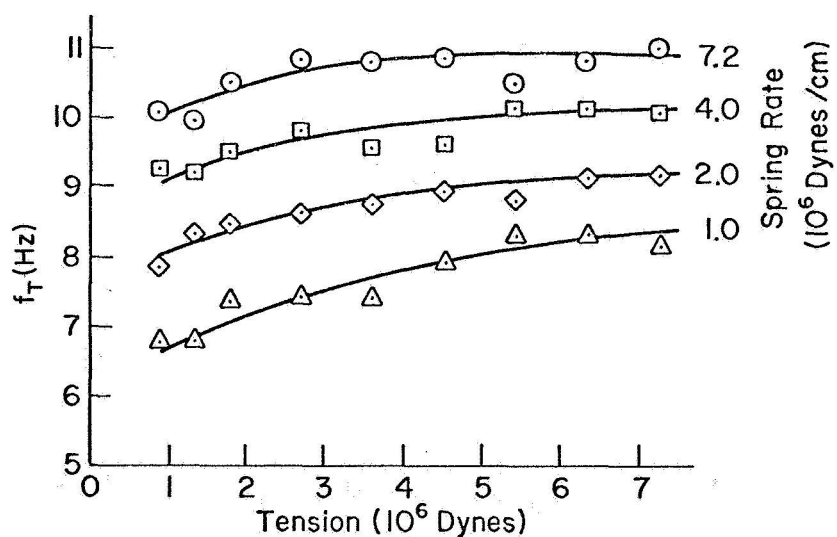


Figure 4. —Effects of muscle tension and spring rate on forearm flexor tremor frequency (adapted from ref. 10).

controlling a small inertia with great pressure. This suggests that the oscillation conditions are strongly dependent on the load, and further that each neuromuscular system is independent of the other.

CLOSED-LOOP NEUROMUSCULAR SYSTEM DYNAMICS

In this section we describe the joint action of the component models (muscle actuation and spindle feedback and actuation subsystems) as they contribute to the neuromuscular system model. This will be used to describe the trends in the overall input/output human operator data just discussed. Our closed-loop neuromuscular system model features spindle feedback of an internal muscle length rather than limb position directly.

NEUROMUSCULAR SYSTEM BLOCK DIAGRAM FOR SMALL PERTURBATIONS.—A combination block and schematic diagram of the muscle/spindle system for small perturbations about an operating point is shown in figure 6 (ref. 2). This illustrates the relationship between the spindle characteristics and the limb/manipulator characteristics. It is to be distinctly understood that the schematic represents perturbation operations about steady-state operating points. Consequently, all the signals indicated vary about zero mean values, and the agonist/antagonist relationships are subsumed in the composite schematic.

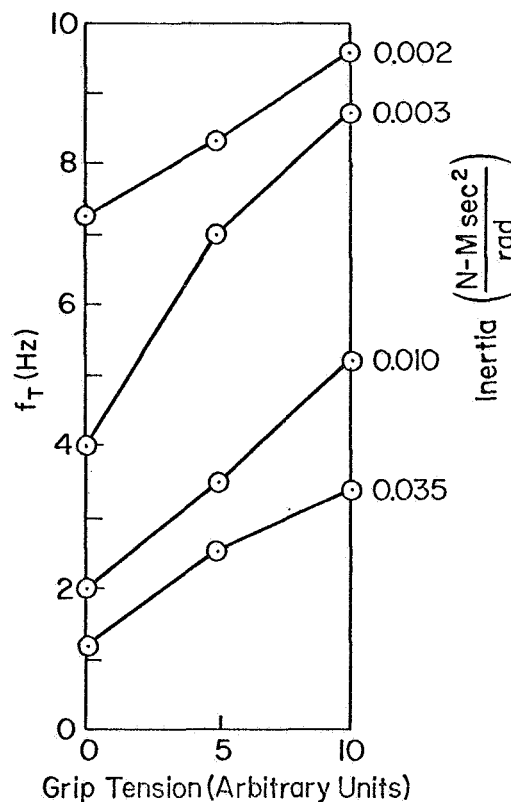


Figure 5. —Effects of inertia and grip tension on limb tremor frequency during constant velocity wrist rotation (ref. 11).

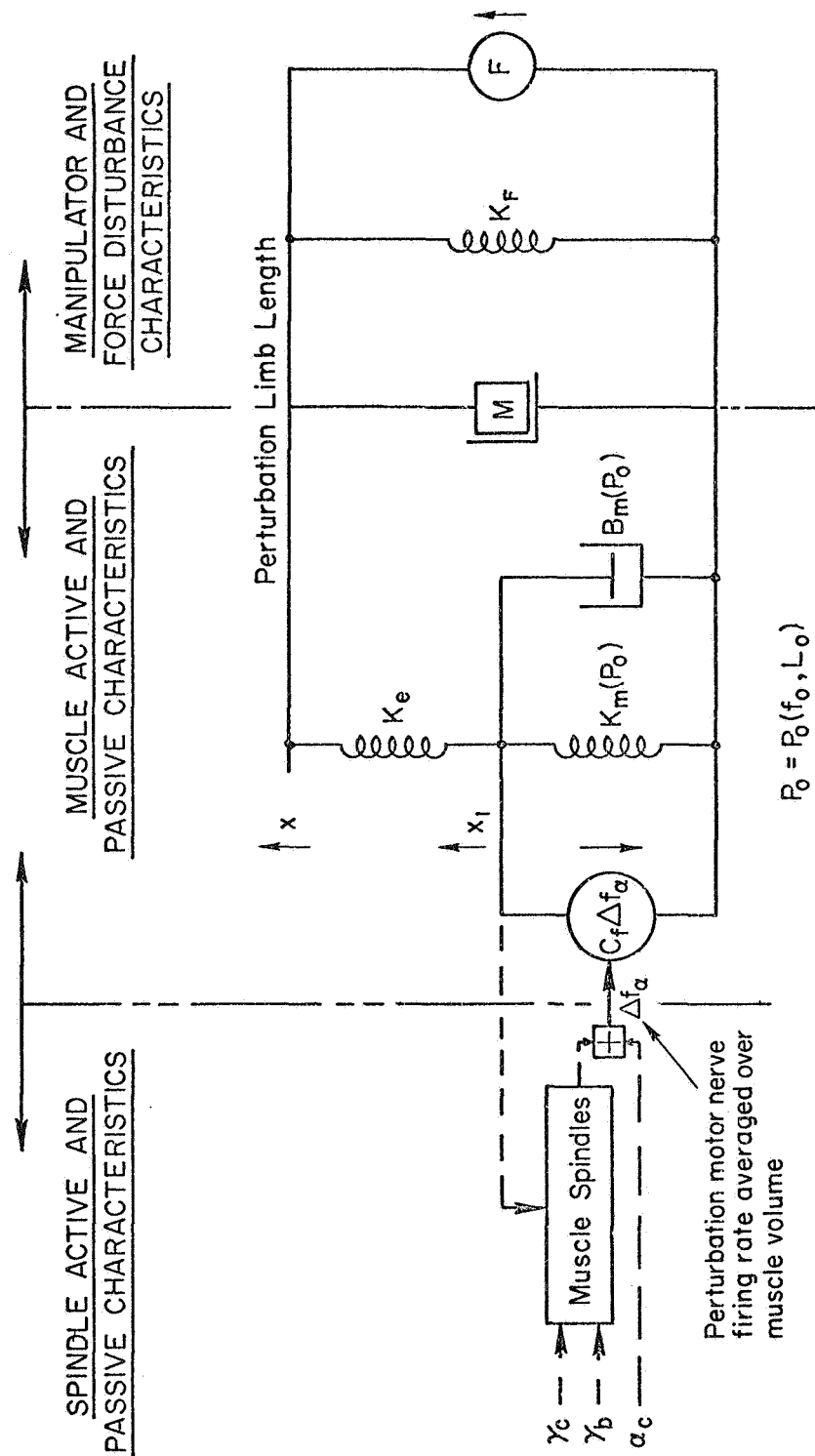


Figure 6. — Combination block and schematic diagram for the muscle/manipulator/muscle spindle system.

The internal inputs to the system are command, γ_c , and bias, γ_b , aspects of the static and dynamic gamma muscle spindle fibers and the alpha motor neuron commands (ref. 2). The external input is a force, F , applied to the manipulator. A key feature is the spindle feedback of x_1 (an internal muscle length) rather than x , the limb length (ref. 2). (Not shown in fig. 6 is a remnant injection quantity which is necessary to account for the system behavior that is not described by the quasi-linear elements.) The muscle model characteristics consist of the series connection of the series elastic component², K_e , with the contractile component characteristics given by the force source,³ $C_f \Delta f_\alpha$, plus spring, K_m , and damper, B_m . Since the limb and manipulator inertias are in parallel, the two are lumped together in the single effective inertia, M . A load spring, K_F , is also present.

The muscle effective damping, B_m , spring rate, K_m , and sensitivity, C_f , are set by the operating point muscle tension, P_o , which is due to the steady-state motor nerve firing rate, f_o , and muscle length, L_o . B_m is a linear function of P_o , whereas C_f and K_m generally increase as P_o increases (ref. 2).

For this paper we do not need to consider the details of the muscle spindle model—only its input/output relations will be needed. The spindle output perturbation firing rate is summed with an alpha motor neuron command, with the result, after conduction and synaptic delays, being an incremental alpha motor neuron firing rate, Δf_α , about the operating point, $f_{\alpha o}$. This in turn perturbs the muscles and manipulator, giving rise to limb rotation, which is sensed by the spindle ensemble. The steady-state output of the muscle spindles (not shown in fig. 6) provides the steady-state muscle firing rate, $f_{\alpha o}$.

We shall first find the open-loop transfer function between the Δf_α and F inputs and the x and x_1 responses. Then we shall apply the spindle feedback dynamics and solve for the closed-loop response.

The equations of motion for the open-loop limb/manipulator dynamics can be found by setting the total force at x and x_1 equal to zero. In terms of Laplace transforms, these can be written as

$$\begin{aligned} F_x = 0 &= -(x - x_1)K_e - (Ms^2 + K_F)x + F \\ F_{x_1} = 0 &= -(x_1 - x)K_e - (B_ms + K_m)x_1 - C_f \Delta f_\alpha \end{aligned} \quad (1)$$

or in matrix form

$$\begin{bmatrix} (Ms^2 + K_F + K_e) & -K_e \\ -K_e & (B_ms + K_m + K_e) \end{bmatrix} \begin{bmatrix} x \\ x_1 \end{bmatrix} = \begin{bmatrix} F \\ C_f \Delta f_\alpha \end{bmatrix} \quad (2)$$

The characteristic equation for the system poles is given by the determinant of the left side of equation (2).

²This is largely due to tendon elasticity—this serves to separate the muscle length that the spindles sense from the actual limb length.

³ Δf_α is a perturbation firing rate about $f_{\alpha o}$. Both of these represent the average behavior over a volume of muscle similar to that which would be measured by an electro-myogram.

$$\Delta = MB_m \left[s^3 + \left(\frac{K_m + K_e}{B_m} \right) s^2 + \left(\frac{K_F + K_e}{M} \right) s + \frac{K_F(K_m + K_e) + K_m K_e}{MB_m} \right] \quad (3)$$

In reference 2 a number of special cases were considered, ranging from an isometric manipulator ($K_F \rightarrow \infty$) to a large inertia, no spring case. It was found that the approximate factors of the third-order characteristic equation indicate that the real root is the same as that for an isometric manipulator. Using this, the approximate factors to equation (3) become

$$\Delta \doteq MB_m \left(s + \frac{K_m + K_e}{B_m} \right) \left[s^2 + \frac{B_m}{M} \left(\frac{K_e}{K_m + K_e} \right)^2 s + \frac{1}{M} \left(K_F + \frac{K_m K_e}{K_m + K_e} \right) \right] \quad (4)$$

which applies as long as

$$\frac{K_m + K_e}{B_m} \gg \frac{B_m}{M} \left(\frac{K_e}{K_m + K_e} \right)^2$$

This approximation holds for large inertia and large K_e . Generally K_e , due to tendon elasticity, is large relative to K_m .

The muscle spindle typically has a lead/lag response to length changes. From figure 6 the perturbation firing rate, Δf_α , depends on the spindle feedback of x_1 and on α_c as (ref. 2).

$$\Delta f_\alpha = [\alpha_c + G_s(\gamma_c - x_1)] e^{-\tau_\alpha s} \quad (5)$$

where $G_s = \kappa_s \left[\frac{s+1/T_K}{s+1/aT_K} \right]$

τ_α = the net effect of synaptic and conduction time delays

γ_c = effective command due to gamma fiber stimulation

$1/T_K$ = the lead break frequency

$1/aT_K$ = the lag break frequency

Now inserting equation (5) into equation (2) and collecting terms that multiply x_1 yields

$$\begin{bmatrix} (Ms^2 + K_F + K_e) & -K_e \\ -K_e & (B_ms + K_m + K_e) + C_F G_s e^{-\tau_\alpha s} \end{bmatrix} \begin{bmatrix} x \\ x_1 \end{bmatrix} = \begin{bmatrix} F \\ (\alpha_c + G_s \gamma_c) e^{-\tau_\alpha s} \end{bmatrix} \quad (6)$$

as the complete system equation for the closed-loop muscle/manipulator/muscle spindle system of figure 6.

The x response to the various forcing function terms F , α_c , and γ_c is

$$x = \frac{[(B_m s + K_m + K_e) + C_f G_s e^{-\tau_\alpha s}] F + [K_e (\alpha_c + G_s \gamma_c) e^{-\tau_\alpha s}]}{\Delta + C_f G_s (M s^2 + K_F + K_e) e^{-\tau_\alpha s}} \quad (7)$$

where Δ is the determinant of the left side of equation (6) in the absence of G_s . When spindle feedback is present, it modifies the closed-loop poles as shown by the denominator of equation (7). However, the spindle feedback, G_s , also modifies the numerator dynamics of the limb response to a force disturbance input. This will have a significant effect on the interpretation of the torque disturbance input data. The effect of the spindle feedback on the limb response to command inputs is primarily through the denominator roots.

In this paper we are interested in two situations:

- (1) An impulse torque disturbance input delivered when the subject's agonist/antagonist muscles are in tension against each other and the manipulator is loaded with a large inertia (no spring restraint). These impulses are delivered at random times and directions, thus the subject's voluntary actions are largely confined to setting the muscle tension so that the resulting response reflects the properties of the x/F transfer function, i.e., α_c and γ_c are zero. The subject's task is merely to set muscle tone and let the transient run its course.
- (2) Limb tremor under various tension conditions where again α_c and γ_c are zero except for ambient noiselike perturbations (also for this case $F=0$).

FORCE DISTURBANCE INPUT (NO SPRING, LARGE INERTIA).—The effect of the spindle feedback on the x/F response can be found by noting that the zeros are given by the roots of the numerator of equation (7) found from the numerator closure as

$$1 + \frac{C_f G_s e^{-\tau_\alpha s}}{B_m s + K_m + K_e} = 0 \quad (8)$$

and the poles are given by the roots of the denominator of equation (7) found from the denominator closure as

$$1 + \frac{C_f G_s (M s^2 + K_F + K_e) e^{-\tau_\alpha s}}{\Delta} = 0 \quad (9)$$

where for this manipulator the roots of Δ are approximated by equation (4) with $K_F=0$ which yields

$$\Delta = M B_m \left(s + \frac{K_m + K_e}{B_m} \right) \left[s^2 + \frac{B_m}{M} \left(\frac{K_e}{K_e + K_m} \right)^2 s + \frac{K_m}{M} \left(\frac{K_e}{K_e + K_m} \right) \right] \quad (10)$$

Both the numerator and denominator of the closed-loop function x/F are functions of the gain parameter $C_f \mu_s$. The joint movement of the numerator and denominator singularities as a function of this loop gain ($C_f \mu_s$) can be demonstrated by treating both the numerator and the

denominator as loop closures. Then both closed-loop poles and closed-loop zeros will move as loop gain is changed. Root locus sketches of the denominator and numerator closures are given in figure 7(a) and figure 7(b), respectively. These two closures have the same real-axis poles and time delay terms (the latter are not shown explicitly since they are at higher frequencies). Only the denominator closure will have quadratic roots (at high frequency). The closed-loop low frequency real root will be approximately the same in both closures. Thus the closed-loop system describing function, x/F , is dominated by a second-order mode which has an undamped natural frequency which is a function of the loop gain (operating point tension dependent) as well as the external mechanical element (inertia). This describes the reference 3 results in figure 2. However, the degree of pole/zero cancellation will be affected by the proximity of the quadratic terms. Some of the different responses are shown in figure 1. In figure 1(b) the response seems to contain some of the first-order pole, whereas in figure 1(a) the response is nearly a pure second-order transient.

Now examine the consequences of assuming that the spindles feed back x instead of x_1 (in fig. 6 let $K_e \rightarrow \infty$). Carrying this through the equations reveals that the numerator of equation (7) will not contain $C_f G_s$; hence the pole/zero near cancellation will not take place. Thus the transient response will be third order. The real root will be more evident at the higher tensions since this increases loop gain and drives the $1/T_K$ root to lower frequency away from the numerator zero. This trend is just the opposite of that in figure 1(b) which becomes more like a second-order system at higher tensions.

TREMOR.—A number of theories have been advanced to account for tremor oscillations, among them that they arise from: (1) periodically fluctuating activity in supraspinal centers, (2) neuronal rhythmicity, (3) resonance of the moving parts of the limbs, and (4) oscillations arising from the stretch-reflex loop itself.

At present, the evidence very much favors the last hypothesis, since: (1) in tabetic patients⁴ with loss of dorsal root fibers, the tremor frequency (6 to 10 hertz) peak is absent; (2) there does not seem to be any correlation of EEG activity (alpha rhythm) with limb tremor; (3) tremor seems to disappear with section of the dorsal roots; and (4) there is no evidence of inherent rhythmicity of spinal alpha or gamma motor neurons in the absence of sensory input. Also, tremor occurs primarily under conditions where fine control of position is involved; and much less tremor is produced under rest conditions or when a gross, rapid (ballistic) type of movement is executed. This also tends to support the stretch-reflex (closed-loop) explanation.

The tremor peak in the frequency spectrum is said by some (e.g., ref. 13) to be indifferent to changes in the load; but the data of references 10, 11, and 12 (figs. 4 and 5) clearly contradict this and, in addition, indicate that the peak frequency is dependent on the mean tension generated, spring rate, and inertia. An excellent example of tremor while tracking (involving fine control of position) is provided by some unpublished data taken during experiments reported in reference 14. This was a multiloop task in which the pilot's feet were controlling rudder pedals and his hand was controlling a spring-restrained side stick. One of the subjects used a high tension technique, i.e., tensed both legs against the pedals and performed the rudder task with differential force changes. Simultaneously, his hand tracking displayed an intermittent tremor of 8 hertz, which was largest at the extreme stick excursions (which requires the largest tensions).

⁴Patients suffering from advanced stages of syphilis who show extreme loss of large sensory afferents, specifically spindle I_a fibers (loss of the stretched reflex) and joint receptors (loss of proprioceptive cues).

The data cited above and earlier indicate that any neuromuscular system model must be compatible with ambient noise excitation of an almost neutrally stable mode near 6 to 11 hertz which varies with inertia, spring rate, and tension. The closed-loop system for these situations in which tremor was measured is described by equation (7), with $F=0$. In addition, α_c and γ_c are nominally zero, although they will exhibit random fluctuations since the human operator cannot generate a perfectly constant tension. These fluctuations will excite the lightly damped closed-loop pole (fig. 7(a)) for the case of inertia restraint and no spring. Increases in inertia will decrease the frequency of this mode, whereas increases in tension, by increasing loop gain, will drive the roots closer to the zeros on the $j\omega$ -axis. Both of these trends are compatible with the data in figure 5.

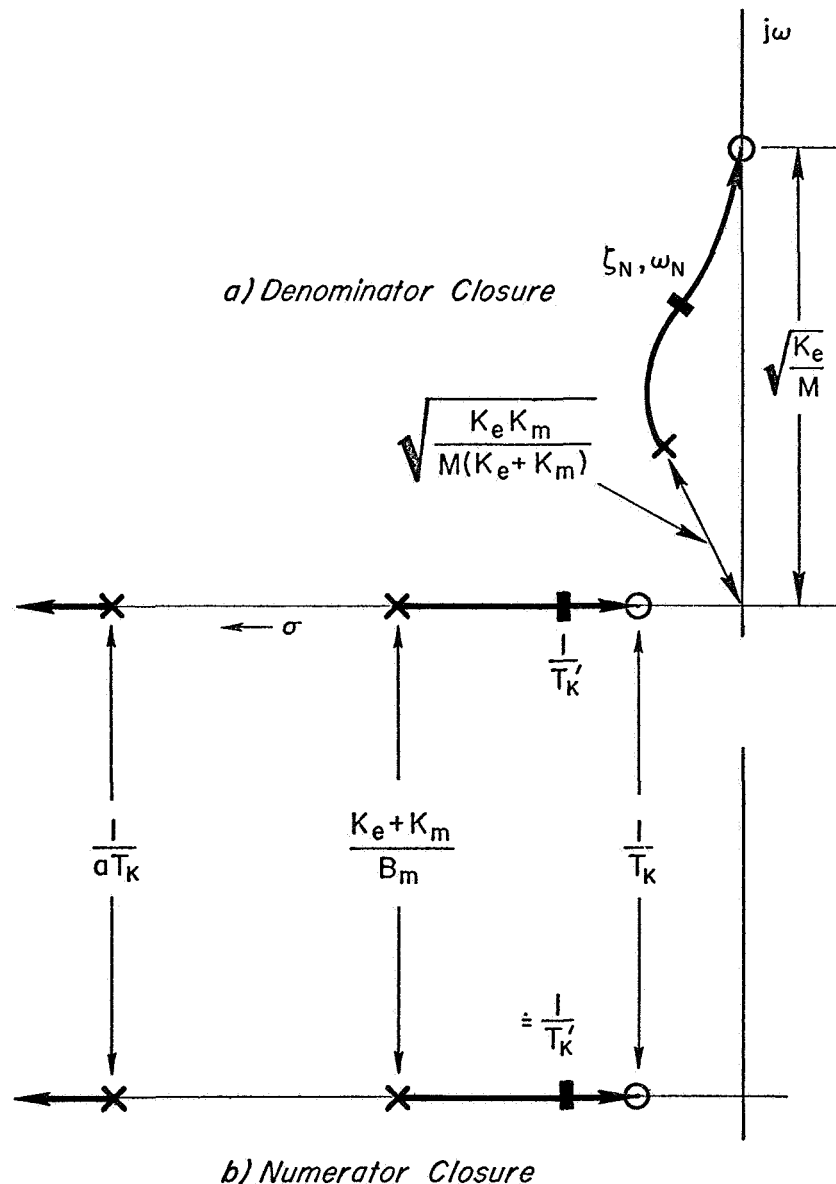


Figure 7.—Root locus sketch for denominator and numerator roots for force disturbance input.

For the data given in figure 4, a spring, K_F , was present. Thus using equation (4) for the approximate roots of Δ and inserting in equation (7) yields the root locus in figure 8. The spring has increased the frequency of the quadratic poles and zeros. Tension increases (by increasing loop gain) drive the mode into the zeros, but now the instability frequency is less than that of the zero due to the phase shift from the time delay $e^{-\tau s}$. The spread between the zero and the instability frequency is likely to be greater for the stiffer spring rates since this raises the zero and the pole to higher frequency where the loop time delay (eq. (7)) has more effect. At higher frequencies the time delay causes the root locus departure from the pole to occur at a lower angle, i. e., it tends to proceed more directly to the $j\omega$ -axis. Thus increases in spring rate yield an instability frequency that is less than the frequency of the zero, and this difference increases at the higher frequencies. This effect

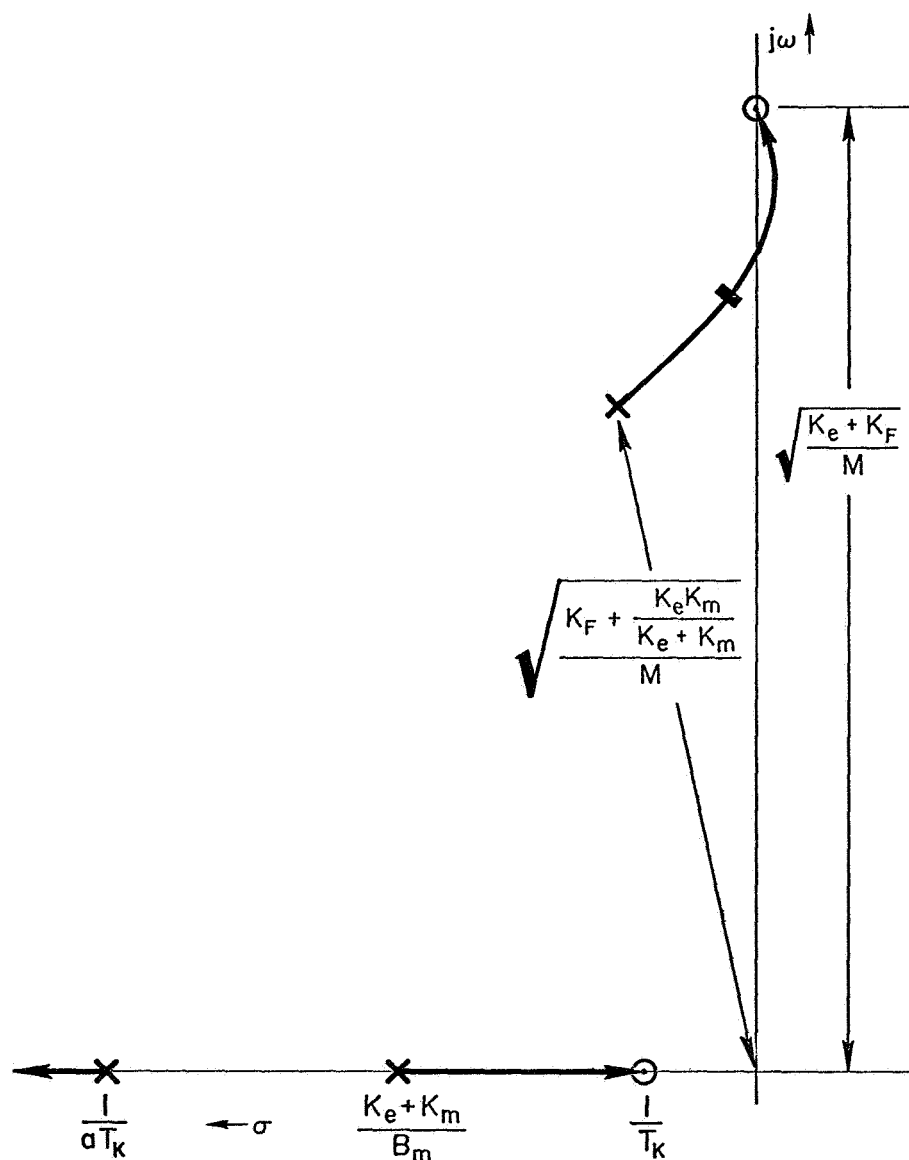


Figure 8. —Root locus sketch for characteristic equation with manipulator spring.

can be seen in the flattening of the curves in figure 4 for the larger spring rates which indicates that the root is near a limit. Attempting to put the above explanation on a quantitative basis by estimating the inertia involved in the figure 4 data, using reasonable values of K_e , and then calculating the effect of K_F yields predicted ω values for the quadratic zeros of between 7 and 14 hertz. These are infinite loop gain tremor frequencies, as compared to the actual tremor frequency range from 7 to 11 hertz given in figure 4.

A time delay on the order of 25 ms is compatible with nerve conduction time from periphery to spinal cord to periphery (refs. 15 and 16). Such a value would contribute 90° phase shift at 10 hertz, leaving 90° to be contributed by the other loop dynamics to produce the 180° instability frequency. The loop gain function (eq. 9) can be compatible with phase shifts of 90° near 10 hertz⁵. Thus the tremor data is compatible with ambient noise excitation of a closed-loop neuromuscular system mode.

Finally the power spectra of finger tremor (ref. 8) exhibited a second peak in the range 17 to 25 hertz, in addition to the usual peak in the range 6 to 10 hertz. This data indicates a higher frequency mode for which a likely source is another lightly damped mode pair due to a real root proceeding in from the time delay singularity at minus infinity combining with the root at $1/aT_K$ (fig. 8) and then nearing neutral stability at frequencies above the quadratic pole/zero combination.

CONCLUSIONS

The data and analysis presented in this report indicate that simple quasi-linear models can describe the small perturbation behavior of the major neuromuscular system components (muscles and muscle spindles) for torque disturbance regulation and limb tremor.

The more specific conclusions are:

(1) Changes in the transient response to a torque impulse input as a function of muscle tension (large inertia, no-spring manipulator) support the model's muscle spindle feedback of an internal muscle length. The effect of this feedback is to produce a closed-loop numerator zero which approximately cancels a denominator root since both are influenced by operating point tension. The resulting dominant second system response closely describes the torque disturbance data.

(2) Variations in muscle tremor frequency for various spring, inertia, and average muscle tension values support a closed-loop neuromuscular system model. The trends and limits of the tremor data provide further evidence of spindle feedback of an internal muscle length. The series elastic element, coupled with limb/manipulator inertias, produces zeros which constrain the closed-loop characteristic equation roots and provide an upper limit for tremor frequency.

REFERENCES

1. McRuer, D. T.; Magdaleno, R. E.; and Moore, G. P.: A Neuromuscular Actuation System Model. Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967, pp. 281-304.

⁵For the special case of an isometric manipulator ($K_F \rightarrow \infty$), the quadratic poles and zeros tend to infinity, leaving only the two real poles and a zero. If the frequency of these singularities is smaller than 10 hertz, then they would contribute approximately 90° phase lag at 10 hertz.

2. Magdaleno, R. E.; Moore, G. P.; and McRuer, D. T.: Small Perturbation Dynamics of the Neuromuscular System in Tracking Tasks. Systems Technology, Inc., Tech. Rept. 154-1, Oct. 1967.
3. Okabe, Y.; Rhodes, H. E.; Stark, L.; and Willis, P. A.: Transient Responses of Human Motor Coordination System. MIT Res. Lab. Elec. Quar. Prog. Rept., no. 66, July 1962, pp. 389-395.
4. Young, Laurence R.; and Stark, Lawrence: Biological Control Systems—A Critical Review and Evaluation, Developments in Manual Control. NASA CR-190, Mar. 1965.
5. Sun, H. H.; Eisenstein, B. A.; and Bomze, H.: Dynamic Model for Hand Motor Coordination System. Eng. in Med. and Biology, Proc. 19th Annual Conference, 1966.
6. Houk, James Charles, Jr.: A Mathematical Model of the Stretch Reflex in Human Muscle Systems. M. S. Thesis, MIT, 1963.
7. Stark, Lawrence: Neurological Feedback Control Systems. Advances in Bioengineering and Instrumentation, ch. 4, Fred Alt, ed., Plenum Press, New York, 1966, pp. 289-385.
8. Wyatt, R. H., Jr.: A Study of Power Spectra Analysis of Normal Finger Tremors. IEEE Trans., vol. BME-15, no. 1, Jan. 1968, pp. 33-45.
9. Holliday, A. M.; and Redfearn, J. W. T.: An Analysis of the Frequencies of Finger Tremor in Healthy Subjects. J. Physiol., vol. 134, 1956, pp. 600-611.
10. Robson, J. G.: The Effect of Loading on the Frequency of Muscle Tremor. J. Physiol., 1959, p. 29P.
11. Gydikov, A.: Sampling with Adjustable Frequency in the Hand Movement Control System. IEEE Trans., vol. HFE-8, no. 2, June 1967, pp. 135-140.
12. Sutton, G. G.; and Sykes, K.: The Variation of Hand Tremor with Force in Healthy Subjects. J. Physiol., vol. 191, 1967, pp. 699-711.
13. Lippold, O. C. J.; Redfearn, J. W. T.; and Vuco, J.: The Rhythmical Activity of Groups of Motor Units in the Voluntary Contraction of Muscle. J. Physiol., vol. 137, 1957, pp. 473-487.
14. Stapleford, R. L.; McRuer, D. T.; and Magdaleno, R.: Pilot Describing Function Measurements in a Multiloop Task. NASA CR-542, Aug. 1966.
15. Margaria, R.; Gualtierotti, T.; and Spinelli, D.: Effect of Stress on Lower Neuron Activity. Exp. Med. Surg., vol. 16, no. 2-3, June-Sept. 1958.
16. Dewhurst, D. J.: Neuromuscular Control System. IEEE Trans., vol. BME-14, no. 3, July 1967, pp. 167-170.

34. A Dynamic Model of an Agonist-Antagonist Muscle Pair*

*W. H. Vickers and T. B. Sheridan
Massachusetts Institute of Technology*

The steady-state force-velocity curve for muscle (both human and isolated) is well established (refs. 1 to 5) to be of the form shown in figure 1. The data are well fitted by a family of hyperbolas:

$$F = zF_0 \frac{1 - V/V_0}{1 + nV/V_0} \quad (1)$$

where z is the percent stimulation (ratio of actual stimulation to maximum; $0 \leq z \leq 1$), V_0 is the maximum velocity, F_0 is the maximum isometric force, and n is a constant.

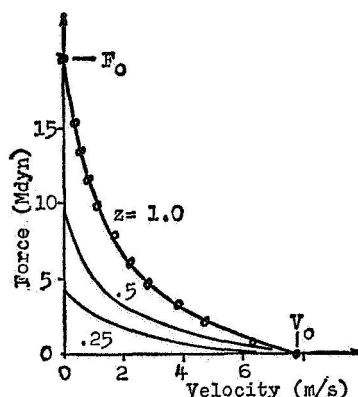


Figure 1. —Muscle force versus steady-state velocity of shortening, both measured at the hand (refs. 2 to 5).

Bigland and Lippold (ref. 5) showed that the stimulation z is proportional to the electromyograph (EMG), at least for steady motion. The force-velocity curve is not as well verified for lengthening muscle, but this relation appears to hold satisfactorily up to the maximum allowable force, which may be twice F_0 .

*This work was supported by NSF and NASA traineeships and Grant NsG 107-61.

The interesting point about this relation is that the stimulation z is multiplicative rather than additive, so that the stimulation, in effect, controls the damping as well as the isometric force. This is a more important nonlinearity than the fact that the lines of constant stimulation are curved in figure 1. In fact, we can linearize this property of (1) to yield

$$F = zF_0[1 - (1 + 1/n)V/V_0]. \quad (2)$$

This F - V relation is valid only for constant velocity; thus to extend the model to cover general motions, a spring must be introduced in series with this steady-state model (fig. 2).

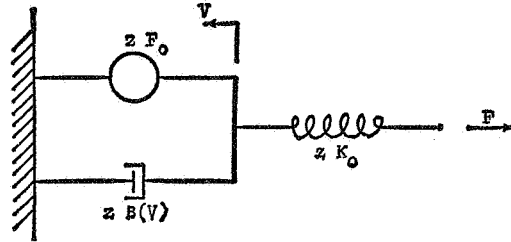


Figure 2.—Schematic model of a single muscle.

If we assume that the effect of increasing the stimulation is to increase the active cross section of the muscle, as in (1), we can relate the spring constant to z . Consider an elastic element with modulus E :

$$F/A = E \frac{\Delta L}{L_0}$$

$$F/\Delta L = K = \frac{EA}{L_0} = \frac{EA_0 z}{L_0}$$

$$K = K_0 z. \quad (3)$$

This result agrees with measurements made by Wilkie (ref. 4) who found the relation between K and the tension level in the muscle shown in figure 3.

We can neglect, for purposes of this model at least, the variation of isometric torque with angle θ . This variation is not large for the forearm because the effects of variation with muscle length tend to be canceled by the changing lever arm.

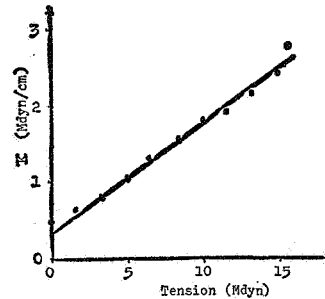


Figure 3.—Effective spring constant of biceps versus muscle tension, both measured at the hand (ref. 5).

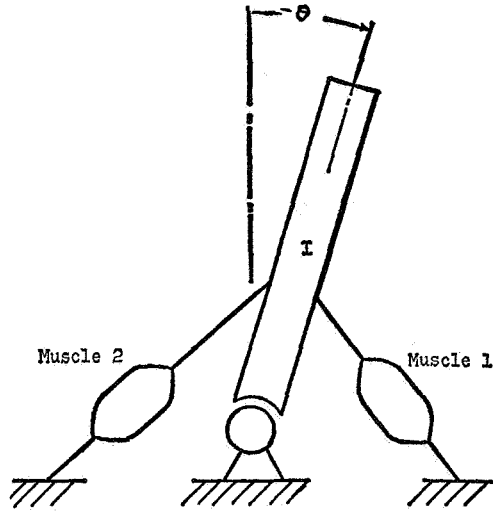


Figure 4.—Schematic of the limb and muscle system.

If we now put two such identical muscles together with a limb (fig. 4), the equations of motion become

$$\begin{aligned}
 V_1 &= l\dot{\theta} + \frac{1}{K_0 l z_1} \dot{T}_1 \\
 V_2 &= -l\dot{\theta} + \frac{1}{K_0 l z_2} \dot{T}_2 \\
 T_1 &= T_0 z_1 (1 - \beta V_1/V_0) \\
 T_2 &= T_0 z_2 (1 - \beta V_2/V_0) \\
 M_d + T_1 - T_2 &= I\ddot{\theta}
 \end{aligned} \tag{4}$$

where l is the lever arm, T_i is the torque exerted by the i th muscle ($T_i = F_i l$), I is the moment of inertia of the limb, the constant $\beta = 1 + 1/n$, and M_d is the disturbance moment. We can nondimensionalize these equations by defining

$$\begin{aligned}
 \mu_i &= T_i/T_0 \\
 v_i &= \beta V_i/V_0 \\
 \omega &= \frac{\beta l}{V_0} \frac{d\theta}{dt} \\
 \tau &= \frac{V_0 l K_0}{\beta T_0} t = \gamma t \\
 \phi &= \int_0^\tau \omega d\tau = \frac{l^2 K_0}{T_0} \theta \\
 R &= \frac{I V_0^2 K_0}{T_0^2 \beta^3} \\
 m_d &= M_d/T_0.
 \end{aligned} \tag{5}$$

Then

$$\begin{aligned}
 \nu_1 &= \omega + \frac{1}{z_1} \frac{d\mu_1}{d\tau} \\
 \nu_2 &= -\omega + \frac{1}{z_2} \frac{d\mu_2}{d\tau} \\
 \mu_1 &= z_1(1 - \nu_1) \\
 \mu_2 &= z_2(1 - \nu_2) \\
 m_d + \mu_1 - \mu_2 &= R \frac{d\omega}{d\tau}
 \end{aligned} \tag{6}$$

Solving for ω ,

$$R \frac{d^2\omega}{d\tau^2} + R \frac{d\omega}{d\tau} + (z_1 + z_2)\omega = (z_1 - z_2) + \frac{dm_d}{d\tau} + m_d. \tag{7}$$

Note that the difference between stimulations ($z_1 - z_2$) appears as a normal linear input, while the sum ($z_1 + z_2$) acts as a parameter. Thus the dynamic properties of the neuromuscular system can be changed simply by changing the level of stimulation.

If we think of ($z_1 + z_2$) as being constant, we can Laplace transform (7). It is convenient to introduce two new variables:

$$\begin{aligned}
 \delta &= z_1 - z_2 \\
 \sigma &= z_1 + z_2.
 \end{aligned}$$

Note that δ and σ are restricted to the range shown in figure 5.

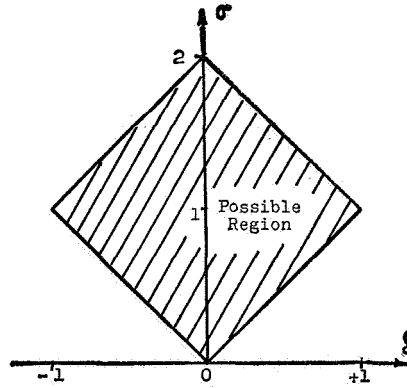


Figure 5.—Constraint relation between σ and δ .

Transforming (7) with Ω , Φ , Δ , and \hat{m}_d representing the Laplace transform of $\omega(\tau)$, $\phi(\tau)$, $\delta(\tau)$, and $m_d(\tau)$, respectively, we obtain

$$\Phi(s) = \frac{1}{s} \Omega(s) = \frac{\Delta(s) + (s+1)\hat{m}_d(s)}{s(Rs^2 + Rs + \sigma)}. \tag{8}$$

Some interesting things appear when we give this system a feedback loop. Assume simply that δ is proportional to the error in position (fig. 6), while σ remains constant. Physically this might correspond to a tracking task or to holding against disturbance forces, where there is likely to be some continuous tension in both muscles.

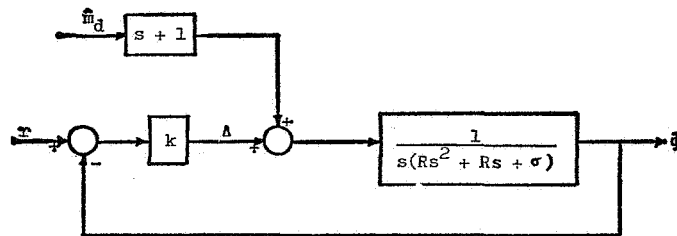


Figure 6. —Block diagram of muscle system with an assumed feedback loop.

To find the effect of tension level σ on the dynamic characteristics, it is helpful to redraw the block diagrams (fig. 7) to make σ appear as the gain in the outer loop. Then to get a locus of roots as a function of σ , we can use the root locus technique on the inner loop first and finally on the outer loop. The inner loop gives a locus like that shown in figure 8. It should be remembered that this is an artificial redrawing of the block diagram; and, therefore, these unstable roots of the inner loop have no real significance for the overall system. The outer loop will stabilize it.

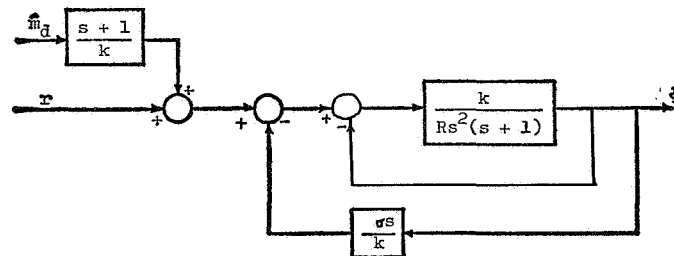


Figure 7. —Block diagram of figure 6 rearranged.

Using these inner roots corresponding to any value of k , we can draw the locus we really want (fig. 9).

The imaginary roots follow Young and Stark's (ref. 6) impulse response experiments amazingly well for $\gamma \approx 22 \text{ s}^{-1}$ and $k/R = 0.054$. From the parameters pertaining to the forearm γ was estimated, while k/R was chosen to give a reasonable root locus. Stark's data, however, were taken for wrist rotation. No attempt was made to adjust for the difference between wrist and forearm dynamics. Stark's data are plotted on the root locus in figure 9. The existence of the first-order pole in the root locus does not detract from the closeness of fit of the model because of the lead term $(s + 1)$ (fig. 7), which is outside the loop and plays a part only when an external torque is present, as was the case for these data. This lead effectively cancels the extra pole (which ranges from about -1 to -0.4) at low tension levels and significantly reduces its importance even at large tension levels. However, if Stark's data are reexamined carefully we might find evidence of this pole for high tension.

This model should be useful in explaining the effects of tension level on the dynamic characteristics of a human controller, as previously discussed. It might also help model the human response to large step inputs. This response frequently appears nonlinear, and

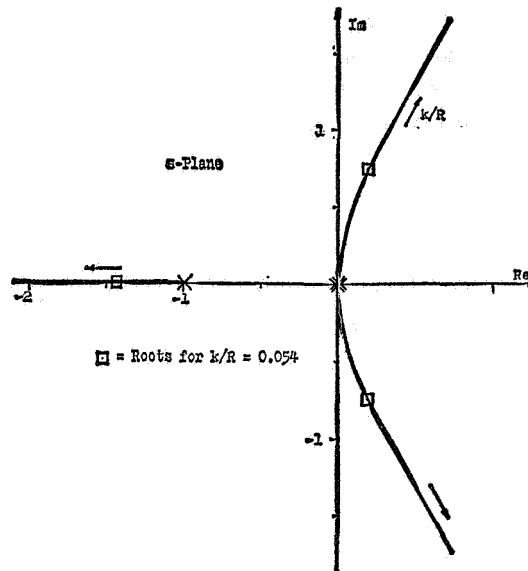


Figure 8.—Root locus for inner loop of figure 7.

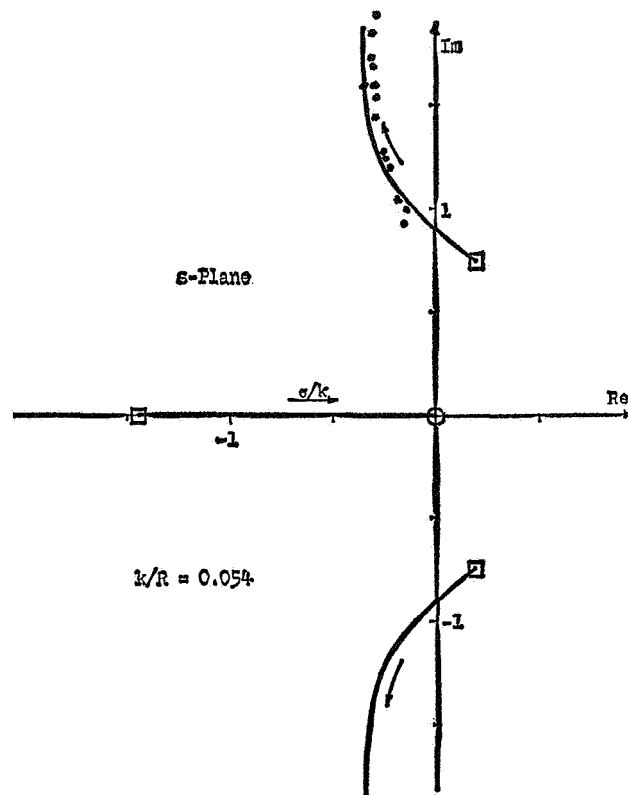


Figure 9.—Root locus for outer loop of figure 7, showing pole locations as a function of tension level σ (ref. 6).

at least one aspect of this nonlinearity (namely, the phenomenon of tensing up at the end of the motion) can be characterized by this model. Also, step responses similar to a wide range of human responses can be generated by this type of model using discrete changes in the two inputs z_1 and z_2 (ref. 7).

Since this model has two independent inputs which can be measured to some degree by the EMG, it might be possible to test the model by using processed EMG data as the input to the model and comparing the output with the actual output. This might also tell us something about the EMG under dynamic conditions.

In addition, the model might have some implications for prostheses that use EMG as inputs. If the parameter changes produced by variation of tension level could be simulated, the limb would have a better "feel." The user could "tense up" for critical maneuvers to get the necessary accuracy.

REFERENCES

1. Hill, A. V.: Heat of Shortening and the Dynamic Constants of Muscle. Proc. Roy. Soc. (London), vol. B-126, 1938, p. 136.
2. Hill, A. V.: A Discussion on Muscular Contraction and Relaxation; Their Physical and Chemical Basis. Proc. Roy. Soc. (London), vol. B-137, 1950, p. 40.
3. Katz, B.: The Relation Between Force and Speed in Muscular Contraction. J. Physiol., vol. 96, 1939, p. 45.
4. Wilkie, D. R.: The Relation Between Force and Velocity in Human Muscle. J. Physiol., vol. 110, 1949, p. 249.
5. Bigland, B.; Lippold, O. C. J.: The Relation Between Force, Velocity, and Integrated Electrical Activity in Human Muscles. J. Physiol., vol. 123, 1954, p. 214.
6. Young, L. R.; and Stark, L.: Biological Control System—A Critical Review and Evaluation. NASA CR-190, March 1965.
7. Vickers, W. H.: Nonlinear Behavior of a Human Neuromuscular Control System. S. M. Thesis, MIT, 1966.

35. Effects of External Loading on Human Motor Reflexes*

*Gyan Agarwal, Bradley Berman, Michael Hogins,
Peter Löhnberg, and Lawrence Stark
University of Illinois*

Achilles tendon reflex in normal human subjects is quantitatively studied from the viewpoint of multiinput, multioutput systems theory. Isometric and isotonic reflex have been elicited in the gastrocnemius-soleus muscle to study the relationship between the initial conditions (muscle tension or length), the stimulus (hammer force) and the reflex response (change in foot torque or angle). In the isometric case, foot-torque response is shown to be strongly related to hammer force, whereas the time at which the maximum foot-torque response occurs is shown to be a function of initial foot torque. In the isotonic case with no external loading, the foot-angle response is strongly related to hammer force, and a plateau in the response is observed. In the isotonic case with external loading, the foot-angle response is strongly related to involuntary initial foot angle as well as the hammer force.

INTRODUCTION

Since World War II, a great deal of attention has been paid to problems of human engineering, particularly in situations where a human operator is part of a process control loop. The analytical treatments of such situations have been made using the concepts of both steady-state and transient analysis and of stochastic process theory (refs. 1 to 5). As the knowledge of neurophysiology has been extended, primarily from animal experimentation under anesthesia, a large measure of success has been achieved in these human engineering studies. The fact remains, however, that a great many rather simple questions about the human motor coordination system remain unanswered.

The elucidation of the spinal reflexes by Sherrington and his followers (ref. 6) undoubtedly forms a major block of information on which our conceptions of the motor system were built. His studies of isolated reflexes beautifully demonstrated the wired-in patterns of reflex response which were present in his decidedly nonphysiologic "spinal cat"

*This work was sponsored in part by ONR (NB0014-67-A-0185), the National Institutes of Health (NB-06197-03, GM-01436, FR-05477), and the Stone Foundation.

preparations. The experimental methods used for his work were designed to stabilize the reflex responses by the exclusions of any cerebral and cerebellar control or modulation of the reflex mechanisms. The medical research done on reflex activity in humans has been primarily from the point of view of clinical diagnosis (refs. 7 and 8).

Achilles tendon reflex has been studied mainly as a diagnostic aid in patients with myxedema, neurosyphilis, hemiplegia and other diseases (refs. 9 to 21). All these investigators have been concerned with measuring half-relaxation time—time between the tap on the tendon and the moment when the muscle has relaxed halfway—and comparing this time in patients and in normal persons. Little or no attention has been given to the amplitudes of the input hammer force and the reflex response. The purpose of this study of Achilles tendon reflex is to carry out a quantitative analysis of the reflex response from the viewpoint of multiinput and multioutput systems theory. Our ultimate goal is to inquire into the nature of the cerebral and cerebellar control of reflex activities by which these seemingly stereotyped responses are molded into a very versatile and adaptive motor coordination system.

EXPERIMENTAL DESIGN

These experiments are designed to demonstrate the variation of myotatic reflex activity with the programming of a task under conditions of identical muscle length and tension. The two tasks are:

(1) The maintenance of a position (muscle length) against a constant or varying force. This is basically a postural type of movement and may be viewed as (a) feedback loop with error detection by the muscle spindle and a constant set point fixed by cerebrum or cerebellum or (b) a feed forward portion of a larger postural system with error detection by inner ear or vision.

(2) The maintenance of a constant force (muscle tension) against a moving or static object. Monitoring in this case should be by golgi tendon organs and/or touch sensors with cerebral processing of the data, including comparison against a reference from memory for error detection, with output directly from motor cortex without the need for a muscle length stabilizing mechanism.

DESCRIPTION OF ANKLE TESTER

The subject was placed in the kneeling position with the lower leg horizontal and a brace anterior to the knee. The sole of the foot was fixed with straps to a plate which was arranged to rotate plantarflexion and dorsiflexion about the axis of the ankle joint. The rotation could be locked in certain anatomical positions for isometric experiments. The Achilles tendon was struck by a hammer which was free to rotate about a fixed axis. A schematic diagram of the apparatus used is shown in figure 1.

The angular position of the footplate was measured via a potentiometer. The force exerted by the subject tending to rotate the footplate was measured by four strain gages attached to the sidearms of the plate and connected in a bridge circuit. The output of the strain gage bridge circuit was amplified with a dc differential amplifier. The force pulse input to the Achilles tendon was measured by two strain gages attached to the hammer. These strain gages were connected in a bridge with two resistors of 330 ohms each, and the output of the bridge was amplified with dc differential amplifier.

The electrical activities (electromyogram - EMG) of the anterior tibial muscle (ATM) and the gastrocnemius-soleus muscle (GM) were obtained using surface electrodes. The

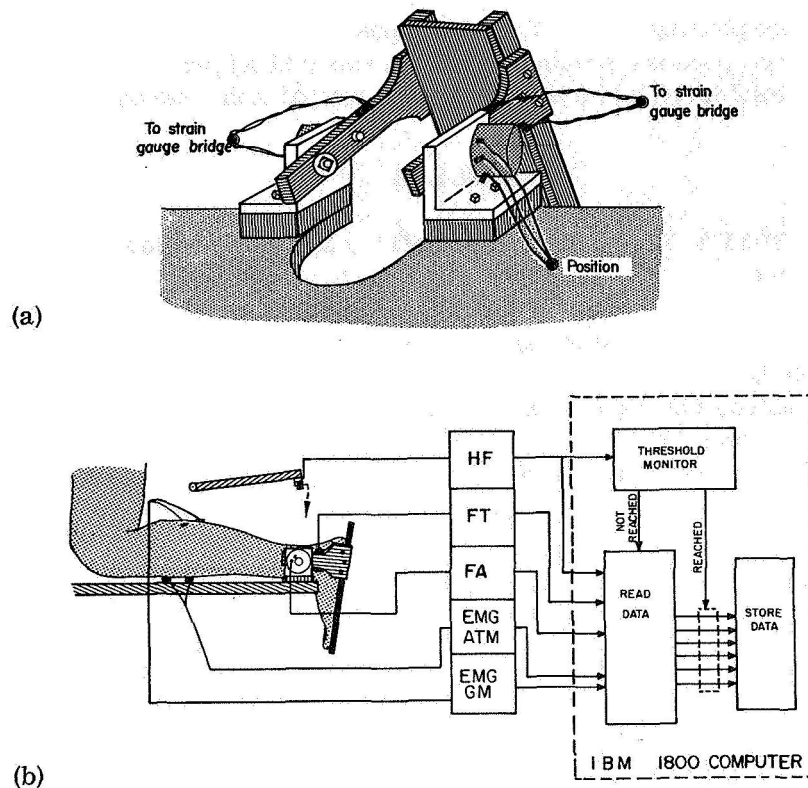


Figure 1. — (a) Schematic diagram of the footplate.
(b) Schematic diagram of the ankle tester.

electrical signal pickups were amplified using ac differential amplifiers. In some experiments, the amplified EMG activity was processed by a high pass filter, a rectifier, and a low pass filter unit as shown in figure 2.

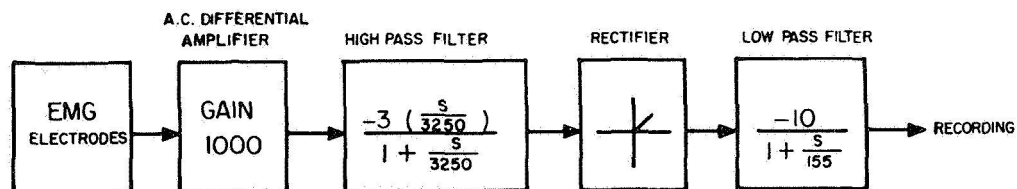


Figure 2. — EMG processing circuit.

The IBM 1800 computer was used for online editing and storage of five channel data on the magnetic tape: hammer force, footplate torque, footplate angle, EMG of anterior tibial muscle, and EMG of gastrocnemius muscle. A special threshold program monitored the hammer output continuously during the course of the experiment; and whenever a preset threshold level was reached, the control was transferred to a data storage program which stored data for 200 msec before and 800 msec after the threshold point at a sampling rate of 1000 per second. The initial foot torque was measured at 100 msec before the hammer pulse input. This program allowed hammer inputs at random intervals to exclude any possibility of the subject's anticipation and also provided for automatic data storage capability. Slowed

reruns of the tests made plotting on an X-Y recorder possible. The computer data processing was done offline. Similar responses were obtained in three normal subjects tested. Most of the data reported here is of one subject (GA) for a more precise comparison between different conditions.

RESULTS

ISOMETRIC CONTRACTION.—In the isometric condition, the footplate is fixed in an anatomical position such that the sole of the foot is at approximately a right angle with respect to the lower leg. The subject is instructed to maintain some constant force in plantarflexion or dorsiflexion direction against the footplate.

In figure 3, the hammer force, foot torque, and EMG activities of the anterior tibial and gastrocnemius-soleus muscles are shown for a typical response. It is interesting to note that the anterior tibial muscle, which is an antagonist, has EMG activity synchronous to the EMG activity in gastrocnemius-soleus muscle. The latency of the EMG activity is about 40 msec from the point of onset of the hammer tap.

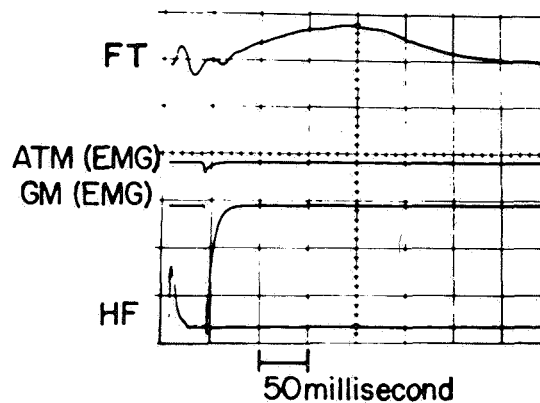


Figure 3.—A typical tendon jerk response in isometric condition.

In figure 4, several foot torque responses for different magnitudes of both the hammer force and the initial foot torque (voluntary contraction) are shown. The shape of the response curves remain unchanged for various magnitudes of these variables. However, the time at which the maximum foot torque response is obtained occurs earlier as the initial foot torque is increased in the plantarflexion direction.

The data analysis is carried out considering the hammer force (HF) and the initial foot torque (IFT) as the independent or the control variables. The dependent or the response variables are (1) the magnitude of maximum change in foot torque in the reflex response with respect to the initial foot torque, this variable will henceforth be called delta foot torque (Δ FT) and (2) the time delay at which this maximum foot torque occurs with respect to the hammer force input, this variable will be denoted by T_m . Due to an initial hammer blow artifact in the foot-torque response, detection of the onset of the response was difficult to program on the computer; and, consequently, this variable, which is roughly equal to 50 msec, was not used for analysis.

The dependence of delta foot torque (Δ FT) on hammer force is portrayed in figure 5. These data clearly suggest a strong functional relationship between Δ FT and HF. The maximum value of Δ FT (0.4 kg-m) is about one-half of the maximum voluntary IFT (0.8 kg-m) used in this experiment. The maximum value of Δ FT is, however, very small (about

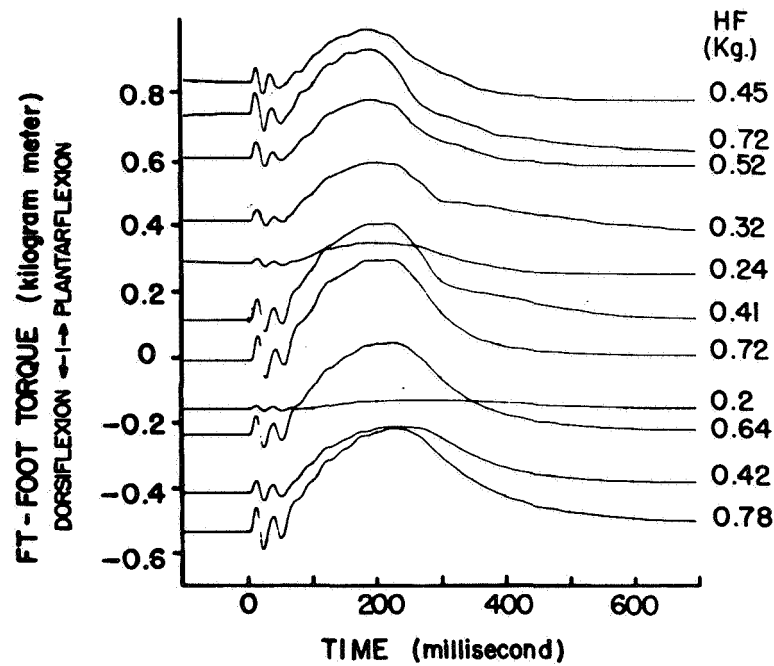


Figure 4. —Foot torque (FT) responses for various magnitudes of hammer force (HF) and initial foot torque (IFT).

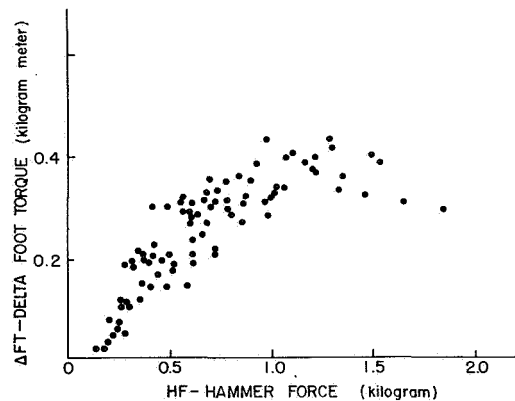


Figure 5. —Delta foot torque (Δ FT) versus hammer force (HF), 0.5 kg-m (dorsiflexion) \leq IFT \leq 0.8 kg-m (plantarflexion).

2 percent) compared to the maximum isometric foot torque that the calf muscles are capable of producing (see discussion). Δ FT decreases slightly for $HF > 1.2$ kg. The initial threshold (dead space) to elicit any measurable Δ FT response is about 0.125 kg, which is approximately 7 percent of input domain of HF.

In figure 6, Δ FT is plotted against IFT (plantarflexion torque considered positive) with HF as parameter. This data is taken for $0.5 < HF < 1.0$ kg. Only modest deviation from a constant function is seen with some falloff at high IFT. The IFT for Δ FT versus HF data in figure 5 range from 0.6 kg-m dorsiflexion to 0.8 kg-m plantarflexion. Small variance in the

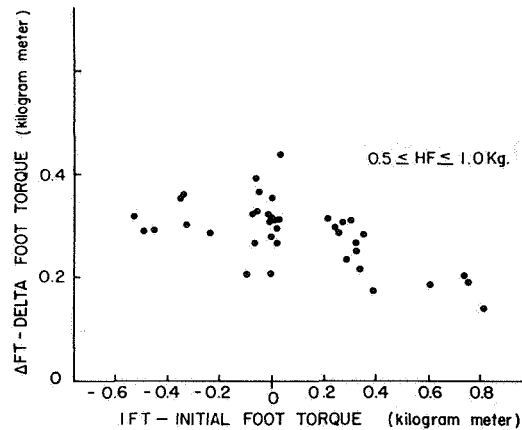


Figure 6.—Delta foot torque (ΔFT) versus initial foot torque (IFT) $0.5 \leq HF \leq 1$ kg.

data, well within the physiological variabilities, suggests additionally that the dependence of ΔFT on IFT is weak as compared to a strong dependence of ΔFT on HF.

The time (T_m) at which the maximum torque occurs is plotted against hammer force in figure 7. These widely scattered data do not suggest any meaningful relation between T_m and HF. In figure 8, T_m is plotted versus initial foot torque (IFT). For IFT in the plantar-flexion direction there is a short latency of about 180 msec; whereas for IFT in dorsiflexion direction the latency is about 220 msec.

In figure 9, the dependent variable, T_m , is plotted against the other dependent variable, ΔFT . This data is for $0.125 < HF < 1.875$ kg and $-0.6 < IFT < 0.8$ kg-m. This plot also does not indicate any meaningful correlation between T_m and ΔFT .

ISOTONIC CONTRACTION.—In the isotonic condition, two different experiments were performed. In the first experiment, the subject was instructed to voluntarily assume various initial foot angles while the footplate was free to rotate without externally applied torque. In the second experiment, the subject was instructed to remain completely relaxed (as far as possible) while the initial foot angle was changed by applying various external torques.

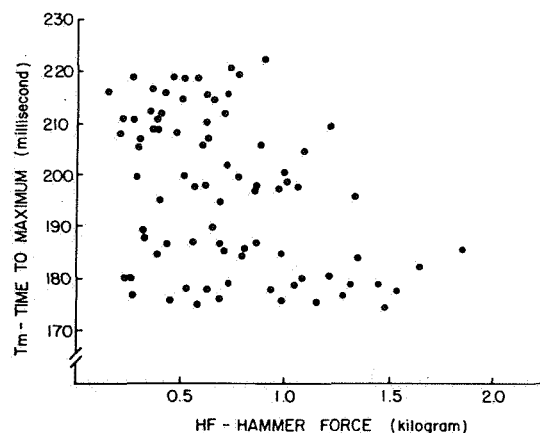


Figure 7.—Time at which the maximum foot torque response occurs (T_m) versus hammer force (HF).

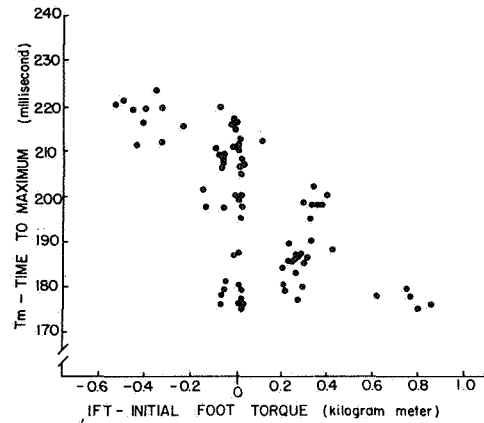


Figure 8. —Tm versus initial foot torque (IFT).

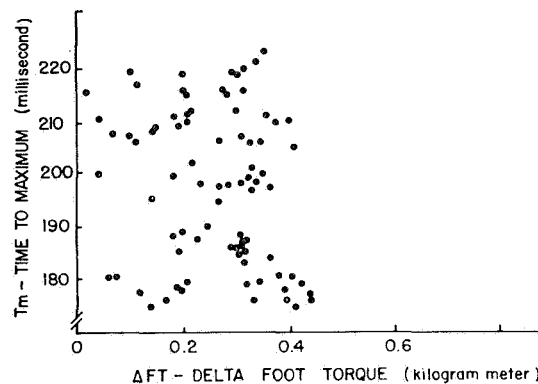


Figure 9. —Tm versus delta foot torque (Δ FT).

Isotonic - Case I.—In figure 10, the hammer force, foot angle, and EMG activities of the anterior tibial and gastrocnemius muscles are shown for a typical response.

In figure 11 (response a), the footplate is at about 90° with respect to the lower leg. In this case, the response starts with a latency of about 60 msec, reaches a maximum value at about 150 msec, remains relatively constant at this maximum value for about 120 msec, and finally returns to the resting position at about 440 msec. In response b, when the foot is in voluntary dorsiflexion, the onset of response is at 60 msec, reaches maximum at 270 msec, remains at the maximum response for 75 msec, and returns to the original position at 650 msec. In response c, when the foot is in voluntary plantarflexion, the maximum is reached at 130 msec, remains at the maximum value for 80 msec, and returns to the base line at 350 msec. The inertia of the footplate is negligibly small compared to the inertia of the foot; and, therefore, it is safe to assume that the external resistive force is zero.

Unlike the foot torque response in the isometric case, the form of the foot-angle response in the isotonic case depends considerably on the initial voluntary foot angle, or, in other words, the initial muscle length. The control variables in this case are the hammer force (HF) and the initial foot angle (IFA). One of the response variables is the magnitude of maximum foot angle in the reflex response with respect to the initial foot angle; this variable will henceforth be called delta foot angle (Δ FA). The second response variable is the time at which the maximum foot angle response occurs; this variable will be denoted by Tm.

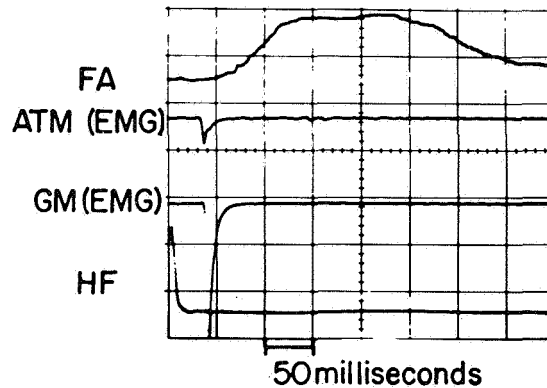


Figure 10. —A typical tendon jerk response in isotonic condition, no external loading.

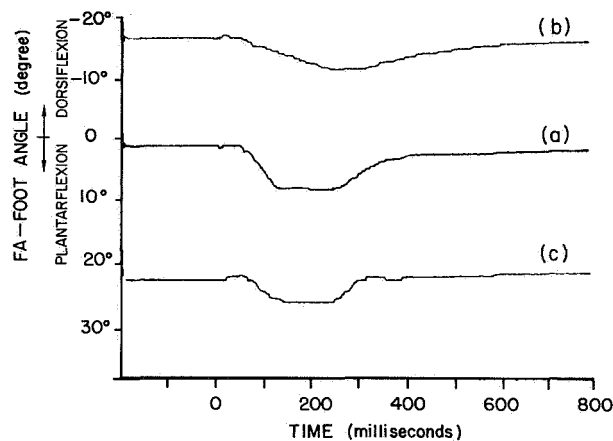


Figure 11. —Foot angle response for various values of hammer force (HF) and voluntary initial foot angle (IFA), no external loading.

The variability of the foot angle response poses some difficulties in the computer analysis of a large number of foot angle responses for ΔFA and T_m . The time at which the maximum ΔFA occurs is not a well defined point. In most responses there is a plateau of duration from about 75 to 120 msec after the initial movement and before the second corrective type response. A small noise signal due to 60 hertz pickup and other sources will cause very little error in the magnitude of FA but may introduce considerable error in defining T_m . Consequently, the variance in T_m is fairly large.

In figure 12, ΔFA is plotted against HF. Here again we notice a strong functional relationship between ΔFA and HF over the range of $0 \leq \Delta FA \leq 8.5^\circ$ for 5° (dorsiflexion) $\leq IFA \leq 10^\circ$ (plantarflexion). The extreme voluntary range for IFA is from about 30° dorsiflexion to about 60° plantarflexion. There is a dead band of about 0.325 kg which is about 16 percent of input domain of HF as compared to 7 percent in isometric case. In figure 13, ΔFA is plotted against voluntary IFA. This data seems to be inadequate to predict any definite relationship between ΔFA and voluntary IFA. In figure 14, T_m is plotted against voluntary IFA. The minimum value of T_m is 145 msec in plantarflexion, whereas the maximum value is 340 msec in dorsiflexion.

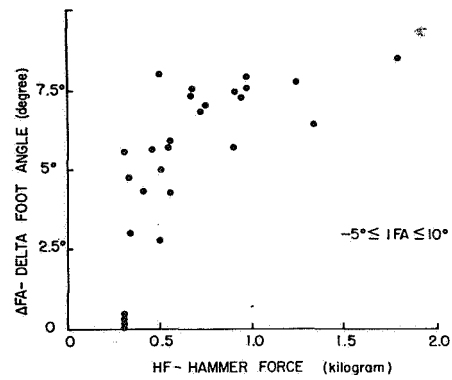


Figure 12. —Delta foot angle (Δ FA) versus hammer force (HF), 5° (dorsiflexion) \leq IFA \leq 10° (plantarflexion), no external loading.

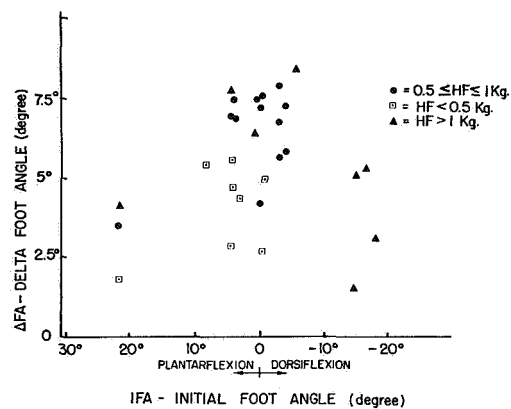


Figure 13. —Delta foot angle (Δ FA) versus voluntary initial foot angle (IFA), no external loading.

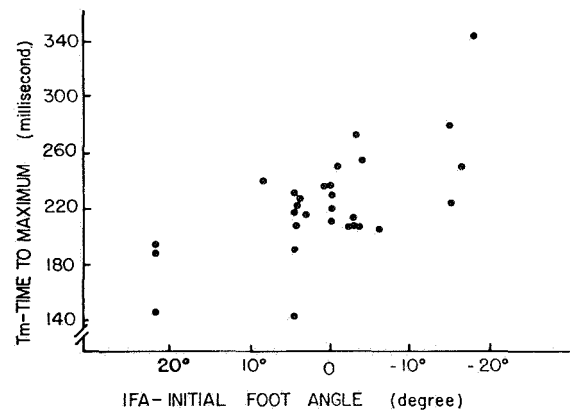


Figure 14. —Time at which the maximum foot angle response occurs (T_m) versus voluntary initial foot angle (IFA), no external loading.

Isotonic-Case II.—While the subject was relaxed, an involuntary initial foot angle was imposed on the foot by means of a torque applied externally to the footplate. A relatively long spring attached to the footplate provides a nearly constant force (torque on the footplate) for small deflections of the footplate; the torque variation in tendon jerk responses amounted to less than 20 percent of the steady state value. The initial involuntary foot angle was varied by changing the external torque (ET), that is, by changing the spring tension.

As the initial involuntary foot angle is varied from extreme dorsiflexion to extreme plantarflexion, the amplitude of the foot angle response decreases. In figure 15, the foot angle response for various initial involuntary foot angles is shown as a function of time. For each of these responses, the hammer force was saturating, that is, it was sufficient to produce a maximal foot-angle deflection. The relationship between delta foot angle (Δ FA) and involuntary initial foot angle (IFA) for a saturating hammer force is shown in figure 16. Note that for initial foot angles of greater than 15° in the plantar direction, Δ FA is essentially equal to zero. The small nonzero values for Δ FA are due to the initial artifact of angular deflection produced by the hammer blow (see last response in fig. 15).

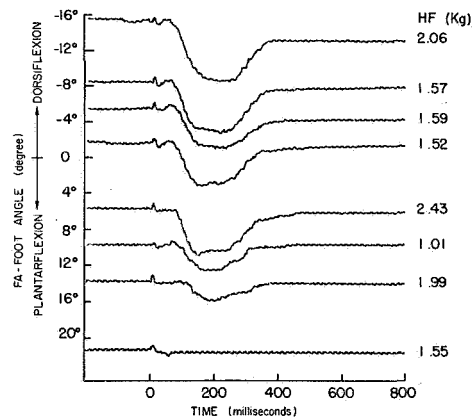


Figure 15.—Foot angle response for various values of involuntary initial foot angle (IFA) due to external torque (ET) and saturating hammer force (HF).

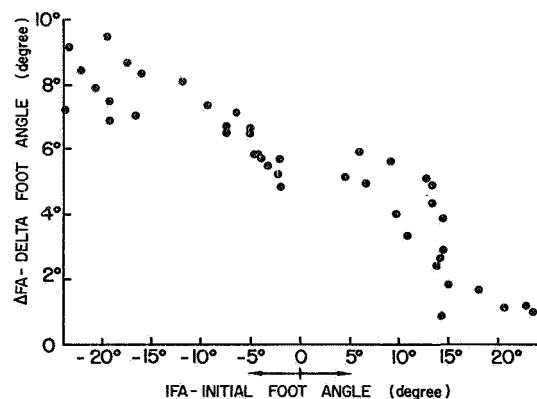


Figure 16.—Delta foot angle (Δ FA) versus involuntary initial foot angle (IFA) due to external torque (ET).

In general, there is a sigmoid relationship between hammer force (HF) and delta foot angle (Δ FA); that is, there is a saturation and initial threshold (dead band). The shape of this sigmoid relationship, however, is a function of the initial involuntary foot angle due to external torque (ET). Δ FT as a function of HF for various initial foot angles is shown in figure 17. As indicated above, the maximum response in figure 17 is observed to decrease as the IFA becomes more plantar. The values of Δ FA below 1° can be attributed to the hammer blow artifact and 60 hertz noise signal. Therefore, from figure 17, the threshold (dead band) to hammer force stimulation increases as the initial foot angle becomes more plantar.

Not only does the amplitude of Δ FA change under different conditions of external loading, but the shape of the foot-angle response curve changes as well. The typical responses shown in figure 15 indicate that under conditions of extreme plantarflexion, the onset of the deflection is increased while the total duration of the deflection is decreased. Under conditions of moderate to extreme dorsiflexion or plantarflexion, the plateau becomes less distinct.

DISCUSSION

ISOMETRIC CONDITION

EMG Activity.—The gastrocnemius muscle is stretched due to the percussion at the tendon. Simultaneously, the anterior tibial muscle is also stretched due to slight movement of the foot and the footplate resulting from the hammer blow. Therefore, it seems that the hammer blow at Achilles tendon stretches both the gastrocnemius-soleus and the anterior tibial muscle nearly simultaneously. This clearly explains the simultaneous EMG activities in the ATM and GM as shown in figure 3 which start with a latency of about 40 msec after the hammer impact. The tendon jerk response, therefore, is not due to the gastrocnemius-soleus muscle alone but is the net response of the gastrocnemius-soleus and anterior tibial muscle acting against each other.

In experiments on the supraspinatus in decerebrate cats, it has been found that a very small stretch stimuli (tap) can elicit a well-marked response. A stretch of 8 microns (circa 0.02 percent) of the tendon is sufficient to elicit a response developing 0.32 kg of tension in this muscle, whose longest fasciculi average some 50 or 70 mm and whose shortest some 26 mm (ref. 22).

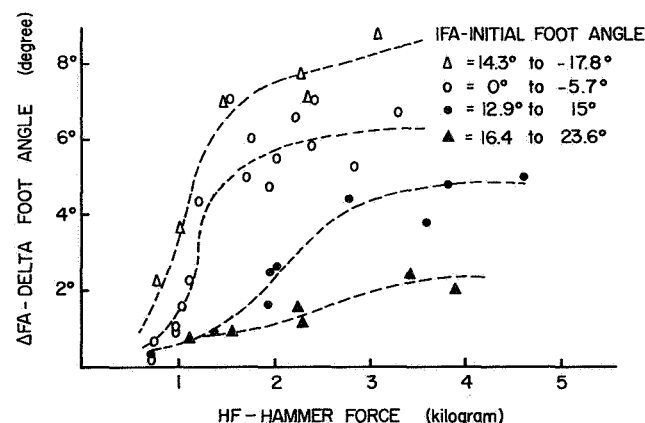


Figure 17.—Delta foot angle (Δ FA) versus hammer force (HF) for various values of involuntary initial foot angle (IFA) due to external torque (ET).

Latency. —The 40-msec time delay in the EMG is comprised of the conduction time to and from the spinal cord, conduction time within the cord itself, and the central delay. In the experiments done by Magladery et al.(ref. 23) on human subjects, the latency of the H-wave in the calf muscle for an electrical stimulus to posterior tibial nerve in popliteal fossa is about 32 msec. The H-wave, as output, corresponds to the mechanically provoked EMG by percussion of the tendon, although the input path is of course more direct (ref. 24). Lambert et al.(ref. 9) have reported a time delay of about 35 msec from start of tap to start of muscle-action potential.

The time delay in the foot-torque response is about 50 msec (fig. 4). This additional 10-msec delay is the time required to develop a measurable force from the moment the efferent volleys reach the muscle. Chaney (ref. 10) has also reported a time delay of about 50 msec between the tendon-jerk stimulus and the beginning of the tension response. The time of onset of the response seems to be independent of the initial foot torque as well as the hammer force.

Δ FT Versus HF. —There was no observable response for hammer force less than 0.125 kg, which, therefore, represents an initial threshold of approximately 7 percent of the input domain of HF (fig. 5). Δ FT initially increases with increased hammer force and attains a peak value of about 0.4 kg-m at about 1.125 kg HF. For HF greater than 1.125 kg, the foot-torque response decreases. The maximum voluntary initial foot torque used in this experiment is 0.8 kg-m. Lippold (ref. 25) measured the maximum tension that could be exerted by the calf muscles of different subjects, ranging from about 30 to 100 kg, with a lever arm of about a foot; that is, the maximum isometric torque ranges from 10 kg-m to 30 kg-m. The maximum Δ FT of 0.4 kg-m in these experiments is only about 2 percent of the maximum isometric torque that the gastrocnemius-soleus muscles are capable of producing.

Granit (ref. 26) has shown that for the soleus stretch reflexes in cats, the reflex tension developed in the muscle is proportional to the amount of stretch (muscle extension). In the present experiments on normal human subjects, any direct measurement of the muscle extension is difficult for both ethical and experimental reasons. The muscle extension or stretch is some nonlinear function of the initial foot torque and the hammer force. The nature of this relationship has not yet been studied. The nonlinear nature of Δ FT versus HF curve (fig. 5) could very well be in part due to the nonlinear nature of the muscle extension as a function of IFT and HF.

Δ FT Versus IFT. —The relationship of Δ FT as a function of IFT with HF as a parameter is shown in figure 6. The range of HF is from 0.5 to 1.0 kg. When IFT is in the dorsiflexion direction, the gastrocnemius-soleus muscle is relatively inactive prior to the hammer input; consequently, the Δ FT response is independent of IFT. For IFT in the plantarflexion direction, Δ FT decreases somewhat as IFT increases. There may be several possible contributing factors for this inhibition when IFT is in the plantarflexion.

The mechanical factor which is in part responsible for this decreased response is a decrease in the spindle receptor stretch when IFT increases. The initial tension in the gastrocnemius-soleus muscle is proportional to IFT (plantarflexion). As the tension in the gastrocnemius-soleus increases, the amount of stretch of the muscle for the same hammer input decreases; and, therefore, the amount of stretch of the spindle receptor decreases. This does not necessarily mean that the afferent signal decreases. The output of the spindle at the same time depends on the gamma inputs and the relative sensitivity of the spindle as a function of the muscle tension.

Another possible explanation of this inhibition comes from the observations by Granit in cat's gastrocnemius muscle (ref. 27). Granit observed that the muscle tension is the

most important factor in determining the form of the excitability curve of the motoneurons under proprioceptive firing. He concluded that the increase of initial tension, within limits, may augment the excitatory phase of the excitability curve, but has fairly regular and relatively greater effect on the inhibitory trough which deepens and expands. In recent studies (refs. 28 to 30) it has been shown that this inhibition is due in part to autogenetic inhibitory afferent impulses from the golgi tendon organs.

Time to Maximum.—The time at which the maximum foot-torque response with respect to the initial foot torque occurs shows a dependence on IFT being less when the initial foot is in plantarflexion than when in dorsiflexion. The range of T_m for the data shown in figures 7 and 8 is from 175 to 223 msec with an average value of about 198 msec and standard deviation of 14.8 msec. The data show a wide variance in the value of T_m within a narrow band about the zero foot-torque position, bridging between a long latency when IFT is in the dorsiflexion direction and a short latency for the plantarflexion direction. The average value from the start of tap to the peak of contraction in normal subjects in the experiments done by Lambert et al. (ref. 9) is 185 msec. Levy (ref. 31) has shown that during a reflex contraction of the calf muscle, the soleus muscle is taking a greater part in this contraction than the gastrocnemius muscle. The gastrocnemius is a fast muscle, having a contraction time in cats in the range of 25 to 40 msec; whereas, the soleus is a slow muscle with a contraction time in cats in the range of 94 to 120 msec (ref. 31, p. 127). Our data in man indicate an average contraction time of about 158 msec from the start of EMG to maximum foot torque. Lambert et al. (ref. 9) have reported the average contraction time in normal man of 150 msec from start of potential to peak of contraction.

The nature of the T_m versus IFT relation may be explained in part by considering the stiffness of the muscle and its relation to the spindle output. When the muscle is tense, as the gastrocnemius is in the case of plantarflexion IFT, a tendon jerk will cause more synchronous output from the spindle than when the muscle is slack. This synchronous afferent signal will generate a synchronous efferent signal which is propagated on α -fibers; consequently, the tension will increase more rapidly. The role of the gamma input and its possible relationship with IFT must also be considered. Also, when the muscle tension is increased, a small change in muscle length due to a hammer blow will be transmitted to the spindle (i. e., change in spindle length) in a relatively shorter time. This time delay, from the moment the hammer strikes the tendon to the time the spindle firing rate increases, is possibly only of the order of a few msec; and consequently, it will contribute only a small amount to this overall spread of about 50 msec. There may possibly be some contribution from the viscoelastic forces in the muscle tending to increase the delay time when the muscle is slack. Figure 7 (T_m versus HF) and figure 9 (T_m versus ΔFT) do not suggest any meaningful correlation between these corresponding variables. The T_m versus IFT relation in figure 8 could possibly be approximated by three piecewise linear portions where the center portion passing through the zero torque position has larger slope.

ISOTONIC CONDITION.

Case I - No External Torque.—In the isotonic response (fig. 10) the EMG activity in the anterior tibial muscle is again due to a small initial stretch of this muscle because of the hammer blow at the Achilles tendon. The foot-angle response, therefore, is the net response of these antagonist muscles (gastrocnemius and anterior tibial).

The plateau in our isotonic responses varies from about 75 to 120 msec. It is interesting to note that this plateau is maximum when the footplate is at about 90° with respect to the lower leg; and the duration decreases both for dorsiflexion and for plantarflexion initial foot angles. The possible origin of this plateau is discussed later on in case II.

In figure 12, ΔFA is plotted against HF for 5° (dorsiflexion) to 10° (plantarflexion) voluntary IFA. There is a strong functional relationship between ΔFA and HF over the range of $0 \leq \Delta FA \leq 8.5^\circ$ for $0.325 \leq HF \leq 2$ kg. The initial dead band (or threshold) is about 0.325 kg HF as compared to 0.125 kg HF in the isometric case or 16 percent of the input domain as compared to 7 percent in isometric case. The ΔFA response is a saturation type function without falloff at large HF (see fig. 5). In figure 13, ΔFA is plotted as a function of voluntary IFA for various ranges of HF. The data is clearly inadequate to define any meaningful correlation between ΔFA and voluntary IFA. The isotonic responses are much more complicated than the isometric responses; and, therefore, large amounts of data and careful analysis is warranted.

In figure 14, the time, T_m , at which the initial maximum foot-angle response occurs (i. e., the onset of the plateau), is plotted against voluntary IFA. There is considerable variation in the data, but it is consistent with the isometric experiment. In dorsiflexion the maximum value of T_m is about 340 msec. This spread of about 200 msec reflects the complexity of these responses.

Case II - With External Torque.—From anatomical considerations, it is seen that when the foot is extremely plantarflexed involuntarily by an external torque, the Achilles tendon is quite slack while the anterior tibial tendon is stretched. When the hammer strikes the slack Achilles tendon, the stimulus of muscle stretch reaches the spindle receptors more attenuated and after a longer latency than if the Achilles tendon is taut, as in involuntary dorsiflexion. This explains, in part, the smaller amplitude of foot-angle response in initial plantarflexion as well as longer latency in response. Also contributing to the decrease of ΔFA under conditions of extreme plantarflexion is the fact that the anterior tibial muscle is being stretched nearly maximally and will allow very little additional stretch.

The plateau in the foot-angle response (with or without external loading) raises some questions as to the origin of this type of response. It appears that the plateau may be caused by an opposing stretch reflex of the antagonist, the anterior tibial muscle. Since the footplate is free to rotate, the anterior tibial muscle is stretched by the gastrocnemius-soleus muscle contracting in the initial phase of its stretch reflex response to the original hammer blow. Although this secondary stretching of the anterior tibial muscle is fairly slow as compared to the quick stretch of gastrocnemius-soleus muscle due to hammer blow, it seems to be sufficiently rapid to reflexly contract the anterior tibial muscle. In the case of no external torque, the EMG activity in the ATM, using surface-type electrodes, was usually observed with a latency of 40 msec after the hammer percussion and synchronous with the EMG activity in GM. On a few occasions a second spike roughly 35 msec after the beginning of the GM contraction was also noted. In case of external loading, this second EMG wave in the ATM is more frequent, particularly in those cases when the velocity of deflection is greatest. It is quite possible that the later part of the EMG activity in the ATM due to its asynchronous nature was not always sufficient to be observed via the surface electrodes. With this difficulty in mind, the secondary EMG activity can be considered as evidence for the opposing ATM stretch reflex.

Since the soleus muscle is a slow muscle and the anterior tibial muscle is a fast muscle, the time courses of their contractions are quite different. The net effect of these two antagonistic muscles is diagrammatically illustrated by the dotted line in figure 18. This dotted line shows a dip which is not seen to be significant empirically since this would involve a lengthening of the plantar muscles at the same time as they are actively contracting, a phenomenon effectively seen only if the lengthening force is quite strong.

Comparison With Other Methods.—Three methods have been reported in the literature for recording tendon reflexes. Chaney (ref. 10) and Lambert et al. (ref. 9) have measured tension essentially in the isometric condition. Time records of the isometric torque in our

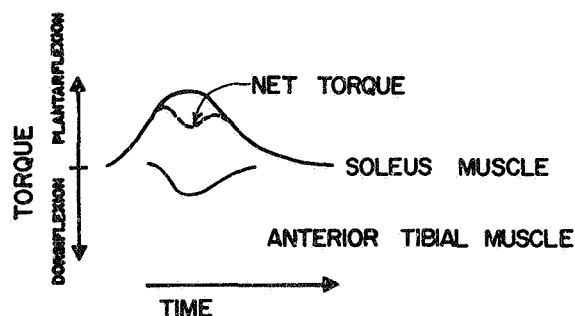


Figure 18.—Interaction of the agonist-antagonistic muscle (soleus and anterior tibial) contractions.

experiments (fig. 4) are very similar to the records given in figure 2 of the Lambert et al. paper (ref. 9). Lawson (ref. 11) has used a kinemometer to measure the free Achilles reflex which is essentially isotonic muscle response. This apparatus used a horseshoe magnet, a coil, and a standard EKG recorder. The voltage recorded is a function of the rate of change of displacement of the foot (i. e., velocity). Gilson (ref. 12) developed an apparatus, the photomotograph which photoelectrically measured the displacement of the foot in timing the free Achilles reflex.

Lawson (ref. 32) has compared the photocell method with the electromagnetic method. He found that the correspondence of values for contraction time showed a coefficient of correlation of less than 0.6, and the corresponding coefficient of correlation for relaxation time was less than 0.5. The linearity of these recording devices and the accuracy which should also include the dynamics of the recorder for these methods have not been reported.

Our method of recording the isotonic response (without external loading) is quite different from the electromagnetic and photocell methods. The resolution and the linearity of the displacement (rotation) measurements depend on the potentiometer used, which is very good (linearity ± 0.25 percent); and the dynamics of the recording device is virtually eliminated since the data is stored on a computer tape at a rate of 1000 samples per second and plotted in slowed reruns.

Smart and Robson (ref. 13) have proposed an apparatus, "oscillomink," which consists of a footplate and a potentiometer device somewhat similar to what we have used for isotonic responses. They observed that their tracings were essentially similar to those reported by Gilson (ref. 12), Sherman et al. (ref. 14), Chaney (ref. 10), and by Lambert et al. (ref. 9) who used other methods. However, no tracings were published in their paper. Since Lambert et al. (ref. 9) only measured the tension in essentially isometric condition, reporting in their paper (ref. 9) that the movement of the platform during the reflex response was less than 0.001 inch, Smart and Robson's observation was certainly incorrect.

SUMMARY

(1) Isometric and isotonic reflexes have been elicited in the gastrocnemius-soleus muscle of normal men in order to study the relation between the initial conditions (tension or length), the stimulus (hammer force), and the reflex response (change in foot torque or angle).

(2) In the isometric case, foot-torque response is found to be directly and strongly related to hammer force, and only secondarily and inversely to initial foot torque. The time at which the maximum foot-torque response occurs is shown to be a function of initial foot

torque and is not dependent on hammer force or the foot-torque response. This time is minimum when the initial tension (IFT) is in the plantarflexion direction.

(3) In the isotonic case with no external loading, the foot-angle response is strongly related to hammer force, showing both a dead band and a saturation. There is a plateau in the foot-angle response which is maximum when the foot-plate is at about 90° with respect to the lower leg. The maximum duration of this plateau is about 120 msec.

(4) In the isotonic case with external loading, the foot-angle response is strongly related to the involuntary initial foot angle due to external torque. The foot-angle response as a function of hammer force shows both a dead band and a saturation.

(5) The plateau in the isotonic responses appears to be caused by an opposing stretch reflex of the anterior tibial muscle; this net effect is also strongly influenced by the differing contractile properties of the agonist-antagonist pair, the soleus and anterior tibial muscles.

REFERENCES

1. Houk, J. C.: A Mathematical Model of the Stretch Reflex in Human Muscle Systems. M. S. Thesis, MIT, 1963.
2. McRuer, D. T.; and Krendel, E. S.: The Human Operator as a Servo-system Element. vol. 267, J. Franklin Inst., 1959, pp. 1-49.
3. Sheridan, T. B.: Human Operator in Control Instrumentation. Progress in Control Engineering - 1. Heywood & Co. Ltd., 1962.
4. Stark, L.: Neurological Feedback Control Systems. Chap. 4 of Advances in Bio-engineering and Instrumentation. F. Alt, ed., Plenum Press, 1966.
5. Young, L. R.; Green, D. M.; Elkind, J. I.; and Kelley, J. A.: The Adaptive Dynamics Response Characteristics of the Human Operator in Simple Manual Control. NASA TN D-2255, 1964.
6. Creed, R. S.; Denny-Brown, D.; Eccles, J. C.; Liddell, E. G. T.; and Sherrington, C. S.: Reflex Activity of the Spinal Cord, Oxford University Press, 1932.
7. Rushworth, G.: Diagnostic Value of the Electromyographic Study of Reflex Activity in Man. Electroencephalography and Clinical Neurophysiology. Supp. 25, Recent Advances in Clinical Neurophysiology. Elsevier Publishing Co., 1967.
8. Magladery, J. W.: Some Observations on Spinal Reflexes in Man. Vol. 261 of Pflügers Archives, 1955, pp. 302-321.
9. Lambert, E. H.; Underdahl, L. O.; Beckett, S.; and Mederos, L. O.: A Study of the Ankle Jerk in Myxedema. J. Clin. Endocrinol., vol. 11, 1951, pp. 1186-1205.
10. Chaney, W. C.: Tendon Reflexes in Myxedema: A Valuable Aid in Diagnosis. JAMA, vol. 82, 1924, pp. 2013-2016.
11. Lawson, J. D.: The Free Achilles Reflex in Hypothyroidism and Hyperthyroidism. New Engl. J. Med., vol. 259, 1958, pp. 761-764.
12. Gilson, W. E.: Achilles-Reflex Recording With a Simple Photomotograph. New Engl. J. Med., vol. 250, 1959, pp. 1027-1028.
13. Smart, G. A.; and Robson, A. M.: A Simple Method for Recording Tendon Reflexes. Lancet, vol. 1, 1963, pp. 363-364.
14. Sherman, L.; Goldberg, M.; and Larson, F. C.: The Achilles Reflex: A Diagnostic Test of Thyroid Dysfunction. Lancet, vol. 1, 1963, pp. 243-245.
15. Fogel, R. L.; Epstein, J. A.; Stopak, J. H.; and Kupperman, H. S.: Achilles Tendon Reflex Test (Photomotogram) as a Measurement of Thyroid Function. N.Y. State J. Med., vol. 62, 1962, pp. 1159-1168.

16. Simpson, G. M.; Blair, J. H.; and Nartowicz, G. R.: Prolonged Achilles Reflex in Neurosyphilis Simulating the Myxedema Reflex. *New Engl. J. Med.*, vol. 268, 1963, pp. 89-91.
17. Simpson, G. M.; Blair, J. H.; and Nartowicz, G. R.: Diagnostic Limitation of Achilles Tendon Reflex in Thyroid Disease. *N. Y. State J. Med.*, vol. 63, 1963, pp. 1148-1153.
18. Buller, A. J.; and Dornhorst, A. C.: The Reinforcement of Tendon-Reflexes. *Lancet*, 1957, pp. 1260-1262.
19. Buller, A. J.: The Ankle-Jerk in Early Hemiplegia. *Lancet*, 1957, pp. 1262-1263.
20. Appenzeller, O.; Imarisio, J. J.; and Gilbert, J. J.: The Effect of Age and Neurological Disease on the Ankle Jerk. *Arch. Neurol.*, vol. 15, 1966, pp. 147-150.
21. Reinfrank, R. F.; Kaufman, R. P.; Westone, J. H.; and Glennon, J. A.: Observations of the Achilles Reflex Test. *JAMA*, vol. 199, 1967, pp. 1-4.
22. Denny-Brown, D. E.; and Liddell, E. G. T.: Observations on the Motor Twitch and on Reflex Inhibition of the Tendon-Jerk of M. Supraspinatus. *J. Physiol.*, vol. 63, 1927, pp. 70-80.
23. Magladery, J. W.; Porter, W. E.; Park, A. M.; and Teasdall, R. D.: Electrophysiological Studies of Nerve and Reflex Activity In Normal Man, IV. The Two Neurone Reflex and Identification Certain Acting Potentials from Spinal Roots and Cord. *Bull. Johns Hopkins Hosp.*, vol. 88, no. 6, June 1951, pp. 499-519.
24. Paillard, J.: Function Organization of Afferent Innervations of Muscle Studied in Man by Monosynaptic Testing. *Am. J. Phys. Med.*, vol. 38, 1959, pp. 239-247.
25. Lippold, O. C. J.: The Relation Between Integrated Action Potential in a Human Muscle and its Isometric Tension. *J. Physiol.*, vol. 117, 1952, pp. 492-499.
26. Granit, R.: Neuromuscular Interaction in Postural Tone of the Cat's Isometric Soleus Muscle. *J. Physiol.*, vol. 143, 1958, pp. 387-402.
27. Granit, R.: Reflex Self-Regulation of Muscle Contraction and Autogenetic Inhibition. *J. Neurophysiol.*, vol. 13, 1950, pp. 351-372.
28. Libet, B.; Feinstein, B.; and Wright, E. W., Jr.: Tendon Afferents in Autogenetic Inhibition in Man. Vol. II of *Electroencephalography and Clinical Neurophysiology*. 1959, pp. 129-139.
29. Hufschmidt, H. J.: Über den Spannungsreflex beim Menschen. Vol. III of *Zeitschrift für Biologie*. 1959, pp. 75-80.
30. Jansen, J. K. S.; and Rudjord, T.: On the Silent Period and Golgi Tendon Organs of the Soleus Muscle of the Cat. *Acta. Physiol. Scand.*, vol. 62, 1964, pp. 364-379.
31. Levy, R.: The Relative Importance of the Gastrocnemius and Soleus Muscles in the Ankle Jerk of Man. *J. Neurol. Neurosurg. Psychiat.*, vol. 26, 1963, pp. 148-150.
32. Lawson, J. D.: Prolonged Achilles Reflex. *New Engl. J. Med.*, vol. 268, 1963, pp. 739-740.

36. Models of Muscle Proprioceptive Receptors*

Gyan Agarwal, Gerald Gottlieb, and Lawrence Stark
University of Illinois

Three lumped parameter models of muscle stretch receptors are described quantitatively on the basis of data obtained from crustacean, amphibian, and mammalian muscle spindles. The first two of these models have single efferent (input) and single afferent (output) innervation. The third model is based on more recent anatomical and physiological studies on mammalian muscle spindles showing two functionally distinct types of efferent innervation and two types of afferent innervation.

INTRODUCTION

In the feedback systems that control animal motor coordination, of fundamental importance is the muscle proprioceptor which detects length and rate of change of length in skeletal muscles. The common type is the muscle spindle, two of which are illustrated somewhat diagrammatically in figure 1 for amphibian and in figure 2 for mammal (refs. 1 and 2).

The polar regions of the spindle consist of striated contractile tissue termed intrafusal muscle. The nuclear bag region, at the spindle's equator, lacks the myofilaments of the intrafusal fiber, is not contractile, and consists mainly of a dense, extracellular network of connective tissue. This region is thought to be stiffer elastically and less viscous than the intrafusal fiber.

The intrafusal muscle is innervated by a diverging and converging set of nerves called the γ -motor system or fusimotor system (ref. 3). Afferent nerve endings of several types are found within the nuclear bag region, the principal one being the annulospiral ending (also called the primary ending).

When the nuclear bag is stretched, either by contraction of the intrafusal muscle or by stretching of the extrafusal fibers, the spindle response is a train of impulses up the afferent nerve, the pulse rate being a function of the degree and rate of stretching.

The purpose of this paper is to show that the important features of this output from a control system viewpoint, may be accounted for on the basis of the mechanical behavior of the spindle, without necessary recourse to more complex molecular and chemical theories.

*This work was sponsored in part by ONR (NB0014-67-A-0185), the National Institutes of Health (NB-06197-03, GM-01436, FR-05477), and the Stone Foundation.

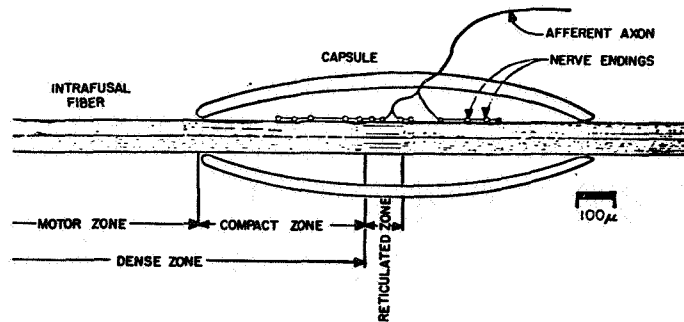


Figure 1.—Diagrammatic representation of an amphibian spindle, after Houk, Cornew, and Stark (ref. 1).

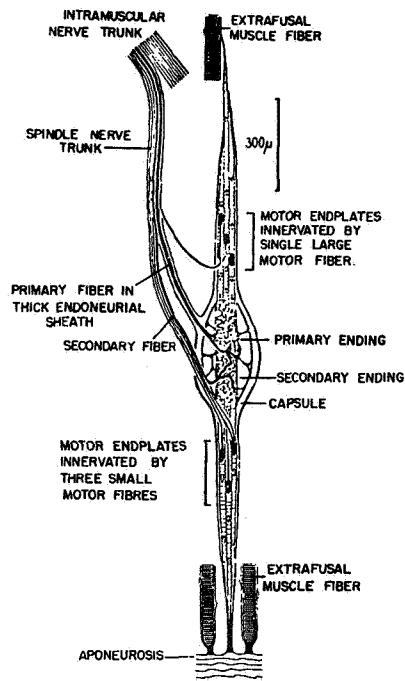


Figure 2.—Diagrammatic representation of a mammalian spindle, after Barker (ref. 2).

Three linear models are considered in this paper. The first model of the amphibian spindle is similar to one suggested by Houk, Cornew, and Stark (ref. 1). This model was developed based on the anatomical findings by Katz (ref. 4), but no attempt was made to demonstrate the validity of this model for the various experimental data known for the amphibian spindle. The second model, of the mammalian spindle, was suggested by Stark (ref. 5). This model is based on an approximation that the primary afferent is the significant output from the spindle, so far as the control of the motor system and the stretch reflex is concerned. There was a slight error in Stark's analysis (ref. 5, p. 303) in developing the transfer function of the spindle. At this time we should like to correct that error as well as show how this new model approximates the experimental data from mammalian spindles for different types of inputs.

The muscle spindle as we have described it thus far is, for reasons of simplicity, deliberately incomplete in that it possesses only one type of γ -motor innervation and one type of output, the Ia primary ending. Recent anatomical and physiological studies by Boyd, Barker, and others (refs. 6, 7, and 8) have clearly shown that the mammalian spindle resembles more closely figure 3 than figure 2. There are two anatomically different kinds of intrafusal muscle fibers within the spindle, named nuclear bag and nuclear chain fibers because of the distribution of nuclei within them. The chain fibers are usually shorter than the bag fibers and are connected to the bag fibers.

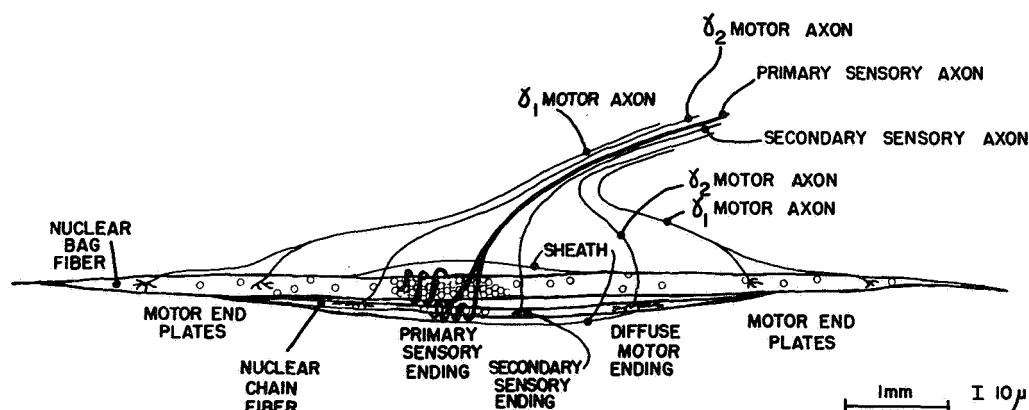


Figure 3.—Diagrammatic representation of a multiple-input-output mammalian spindle, after Boyd (ref. 11).

For each spindle there is at least one primary sensory nerve which makes synaptic contacts around the bag portion of the nuclear bag fiber and often has additional branches to the equatorial regions of the chain fibers. There are also usually one or more secondary endings which synapse almost exclusively with the chain fibers away from the equatorial region (ref. 9). (It is hardly fair to describe these endings as newly discovered since Ruffini in 1898 (ref. 10) described flower-spray endings but only recently has the more detailed anatomy or any physiology been explored.)

There are also two types of motor innervation, called by Boyd γ_1 and γ_2 motor inputs (refs. 6 and 11). Type γ_1 makes contact with motor endplates at the poles of the bag fibers while type γ_2 connects with fine, diffuse networks of fibrils on the chain fibers. These networks are usually polar to the sensory endings but sometimes overlap with the secondary endings.

Corresponding to these two anatomically distinct fusimotor inputs, two distinct types of spindle response to γ stimulation have been found. These have resulted in a physiological classification of fusimotor nerves into dynamic and static categories depending on the nature of the spindle response (ref. 8). This distinction will be considered in further detail later in the discussion.

The third model of the muscle spindle has both nuclear bag and chain components as well as independent fusimotor control over each. Our analysis at this stage is grossly simplified but serves to illustrate several important behavioral characteristics of the multiple-input-output spindle. A qualitative model of cat muscle spindle has been proposed by Andersson et al. (ref. 12).

DESCRIPTION OF MODELS

AMPHIBIAN SPINDLE.—Figure 1 shows the anatomical structure of the amphibian spindle (ref. 4). We propose the mechanical model of figure 4 for this spindle.

The important features of this model are that the reticulated zone of the spindle is assumed to be a purely elastic linear element while the dense zone is taken as primarily a viscous linear element in parallel with a weak spring and a force generator. This force component is the contractile element of the intrafusal muscle.

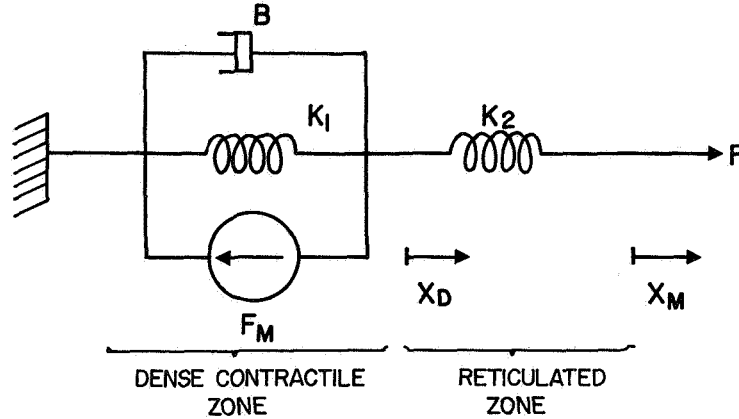


Figure 4. —Lumped parameter model of the amphibian spindle.

The rate of afferent firing f in pulses per second (pps) is assumed to be proportional to the receptor potential developed in the dendritic structure of the primary sensory nerve, and this in turn is assumed linearly proportional to the strain of the dense zone X_D/X_{D0} , the reticulated zone X_R/X_{R0} (X_{D0} and X_{R0} are the rest lengths of the dense and reticulated zones, respectively) and the number of dendritic ending N_D and N_R in the respective zones (ref. 1). This relationship is given by the equation

$$f = k_f \left[N_D \frac{X_D}{X_{D0}} + N_R \frac{X_R}{X_{R0}} \right] \quad (1)$$

where k_f is a constant of proportionality.

Using the Laplace transform (ref. 13), a transfer function for the spindle may be written as in equation (2) (app. A).

$$f = K_X \frac{T_2 s + 1}{T_0 s + 1} X_M + K_F \frac{1}{T_0 s + 1} F_M \quad (2)$$

where

$$K_X = \frac{k_f N_D}{X_{D0}} \left[\frac{K_1 + K_2}{K_1 + K_2} \right]$$

$$K_F = \frac{k_f N_D}{X_{D0}} \left[\frac{-1}{K_1 + K_2} \right]$$

$$T_2 = \frac{\propto B}{\propto K_1 + K_2}$$

$$T_o = \frac{B}{K_1 + K_2}$$

$$\propto = \frac{N_R/X_{Ro}}{N_D/X_{Do}}$$

Equation (2) gives the total change in the frequency of firing f due to a change in the length of the intrafusal fibers caused by the change in the length of extrafusal fibers and a change in the intrafusal force resulting from variation in the γ -efferent activity.

SINGLE INPUT-OUTPUT MAMMALIAN SPINDLE.—A second, more general model, for the mammalian spindle of figure 2 has been proposed on the basis of the known behavior of physical systems (ref. 5). This model is shown in figure 5.

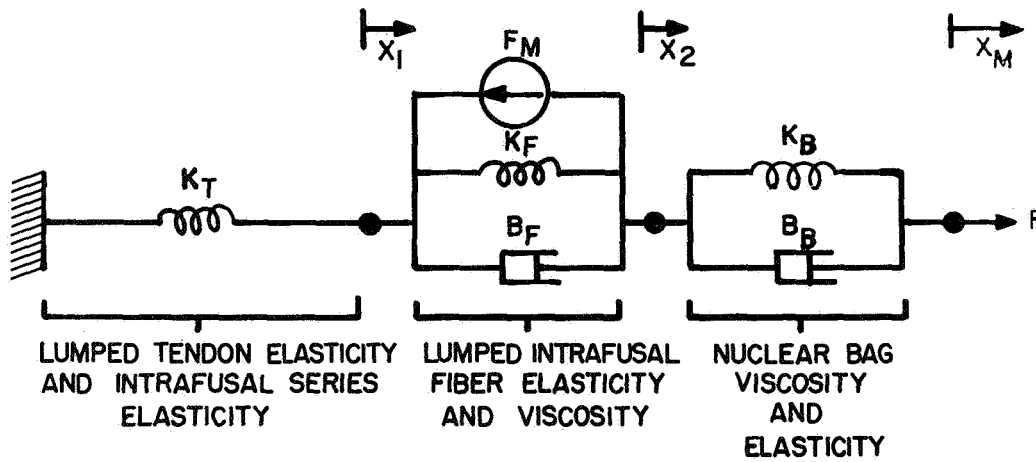


Figure 5.—Single input-output lumped parameter model of the mammalian spindle.

Here, the tendon is lumped with the series elastic elements of total value K_T . The intrafusal muscle has both elasticity, viscosity and a contractile component, and the nuclear bag is characterized by parallel elastic and viscous elements. Since for mammalian muscle only changes in the length of the nuclear bag affect the output,¹ the firing rate in the afferent nerve is proportional to $X_M - X_2 = X$ for which the transfer function is given in equation (3) (app. B).

$$X = K_1 \frac{s + a}{(s + \frac{1}{T_1})(s + \frac{1}{T_2})} X_M + K_2 \frac{1}{(s + \frac{1}{T_1})(s + \frac{1}{T_2})} F_M \quad (3)$$

¹Actually stretch of the nuclear bag provides only one component of the primary afferent response. The second component due to stretch of nuclear chain fibers is neglected here.

where

$$K_1 = K_T/B_B$$

$$K_2 = K_T/B_B B_F$$

$$a = K_F/B_F$$

$$\frac{1}{T_1} + \frac{1}{T_2} = \frac{K_T B_F + K_B B_F + K_F B_B + K_T B_B}{B_B B_F}$$

$$\frac{1}{T_1 T_2} = \frac{K_T K_F + K_B K_F + K_B K_T}{B_B B_F}$$

MULTIPLE INPUT-OUTPUT MAMMALIAN SPINDLE.—A simple, preliminary multiple-input-output mammalian spindle model is shown in figure 6. In it, we have made a number of critical assumptions, some of which seem reasonable enough to keep until and unless they are shown to be inadequate or erroneous and some of which are for the sake of simplicity and will be omitted from future, more sophisticated versions of this model.

First, we are characterizing the contractile portion of the nuclear bag fiber as purely viscous (the force generators which would be in parallel with B_1 , B_2 , and B_3 have been omitted) and the nuclear-bag region as purely elastic.

Second, we have assumed that the nuclear chain fiber is fairly homogeneous over its length, differences between equatorial (B_4) and polar (B_3) regions being quantitative rather than qualitative.

Third, although the evidence in favor is only circumstantial, we identify the anatomically defined type γ_1 motor fibers with the physiologically defined dynamic motor fibers, and similarly, the type γ_2 motor fibers with the static motor fibers (ref. 8, Bessou and Laporte, p. 81).

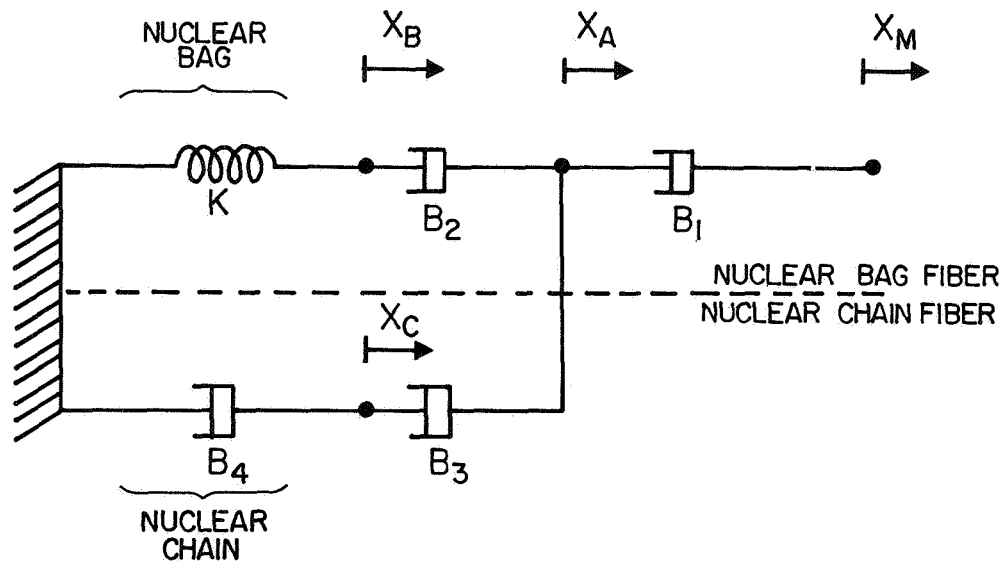


Figure 6.—Multiple input-output lumped parameter model of the mammalian spindle.

Furthermore, since the force generators have been omitted from this model, the only effect of fusimotor stimulation is an assumed increase in the viscosity of the activated muscle fibers. Thus, dynamic fiber stimulation causes B_1 and B_2 to increase while static stimulation causes B_3 to increase.

The transfer functions of this simplified model are given by equations (4), (5), (6), and (7) (app. C).

$$f_p = [K_B T_B(s) + K_C T_C(s)] X_M(s) \quad (4)$$

$$f_s = K_{CS} T_C(s) X_M(s) \quad (5)$$

$$T_B(s) = G_B \frac{s}{s + \frac{1}{T_o}} = \frac{X_B(s)}{X_M(s)} \quad \left| \quad X_C(s) = 0 \right. \quad (6)$$

$$T_C(s) = G_C \frac{s + \frac{1}{T_1}}{s + \frac{1}{T_o}} = \frac{X_C(s)}{X_M(s)} \quad \left| \quad X_B(s) = 0 \right. \quad (7)$$

where

$$\frac{1}{T_o} = \frac{(B_1 + B_2)(B_3 + B_4) + B_3 B_4}{(B_1 + B_3) B_2 B_4} \cdot K$$

$$\frac{1}{T_1} = \frac{K}{B_2}$$

$$G_B = \frac{B_1 (B_3 + B_4)}{B_4 (B_2 + B_3)}$$

$$G_C = \frac{B_1 B_3}{B_4 (B_2 + B_3)}$$

The primary response f_p is proportional to the amount of stretch of both bag $T_B(s)$ and chain $T_C(s)$; K_B and K_C are proportionality constants having units of pulses per sec/mm of spindle stretch. The secondary response f_s is proportional to the amount of stretch of the nuclear chain fibers along $T_C(s)$; K_{CS} is again a constant of proportionality having units of pps/mm.

SYNTHESIS AND ANALYSIS

These models have been synthesized on a digital computer. As a matter of historical fact, we first computed the parameters of the single input-output mammalian spindle model transfer function by minimizing the RMS error for the first 200 msec of the step response based on data from Lippold, Nicholls, and Redfearn (ref. 14). For the amphibian model, we then chose its time constant equal to the principal mammalian time constant of 60 msec. This differs considerably from the 5-second value mentioned by Houk et al. (ref. 1). Since we expected that the viscosity of the intrafusal muscle dominates its behavior, we let

$K_1=0$. (Whether this assumption is valid depends upon the relative strength of the viscous and elastic forces which will be considered in the discussion. The success of our simulations seems to justify it.)

AMPHIBIAN SPINDLE.—From Houk et al. (ref. 1) we got estimates of α ranging from 6 to 8.5 and chose $\alpha=7$ for our simulation. The gain constant K_X was chosen to give a good fit by eye to the experimental data shown in figure 7. This figure shows the amphibian model response to four constant stretch velocity inputs. The discrete points plotted in figure 7 are from experiments performed on isolated spindles of musculus extensor digitorum longus - IV of the frog (ref. 15). In these experiments, one end of the spindle was held in a fixed clamp, while at $t=0$ the other end was stretched at constant velocity from some rest length to an additional 0.35 millimeter (estimated from curves in (ref. 15)). In figures 7(a), (b), (c), and (d), the velocities are 13, 2.6, 1.2, and 0.7 mm/sec, respectively. The continuous curves are the model response at these stretch velocities.

The continuous curve of figure 8 shows the response of the model to four successive, discrete pulses of force generated by the intrafusal muscle F_M . The discrete points show the afferent response of tenuissimus muscle spindles in the cat to four discrete shocks administered to the γ -motor nerve (ref. 16). The model response is normalized to the experimental data because a numerical relationship between γ -efferent pulses and intrafusal force is not known.

SINGLE INPUT-OUTPUT MAMMALIAN SPINDLE.—The behavior of the mammalian model is shown in figures 9, 10, and 11. Figure 9 shows the response to a quick stretch with experimental data from cat's tenuissimus muscle (ref. 14). It is on the basis of these

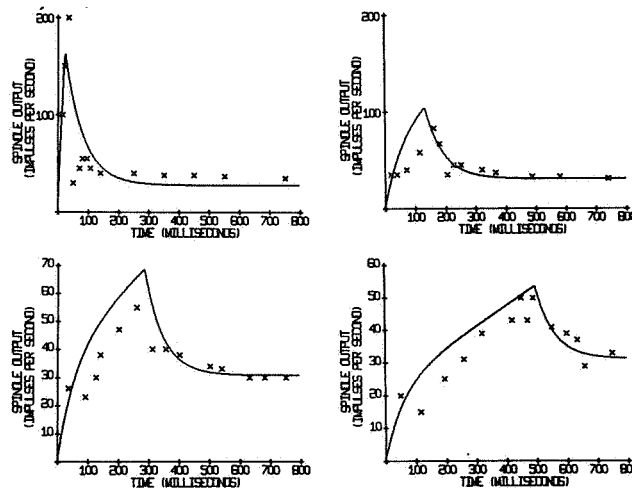


Figure 7.—Response of the amphibian spindle to constant velocity stretch. The continuous curves are the model response. The discrete points are redrawn from the experimental data of Shepherd and Ottoson (ref. 15). The stretch velocities are (a) 13 mm/sec, (b) 2.6 mm/sec, (c) 1.2 mm/sec, and (d) 0.7 mm/sec.

data that the parameters are set at $T_1=0.0055$ second, $T_2=0.06$ second, and $a=3.55$ second $^{-1}$. The transfer function is given by equations

$$\left. \begin{aligned} \frac{X}{X_M} &= \frac{83 (s + 3.55)}{(s + 182)(s + 16.7)} = T_X \\ F_M &= 0 \end{aligned} \right\} \quad (8)$$

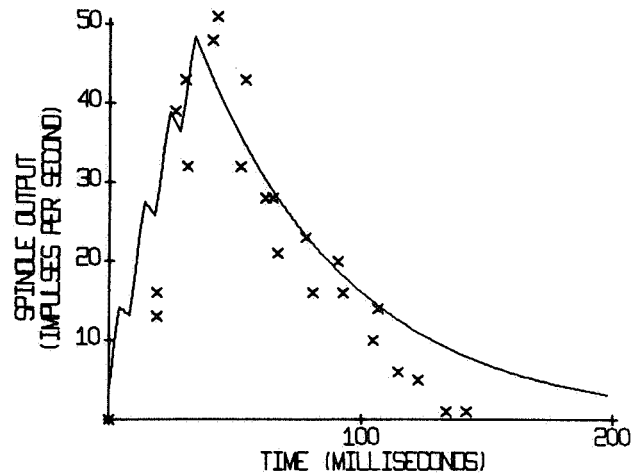


Figure 8. —Response of the amphibian spindle to four intrafusal muscle twitches, γ -efferent input. For the model response (continuous curve), the input is a series of four rectangular pulses of force F_m , each 5 msec wide, occurring at time $t=0, 10, 20, 30$, msec, respectively. The discrete points are the experimental values taken from Eyzaguirre (ref. 16) and resulting from four discrete electrical shocks applied to γ -motor nerve in cat's tenuissimus muscle spindle.

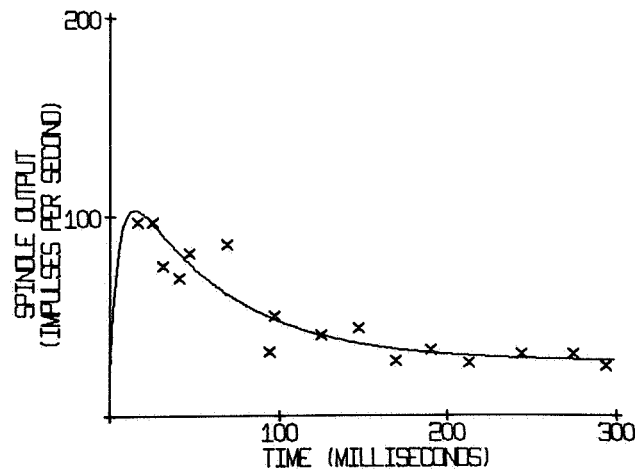


Figure 9. —Step response of the mammalian spindle. The continuous curve is the model response. The discrete points are taken from the experimental data of Lippold, Nicholls and Redfearn (ref. 14).

$$\left. \begin{aligned} \frac{X}{F_M} &= \frac{656}{(s + 182)(s + 16.7)} = T_F \\ X_M &= 0 \end{aligned} \right| \quad (9)$$

$$X = T_X X_M + T_F F_M \quad (10)$$

Figure 10 shows the response of the mammalian spindle model to the same four discrete pulses used on the amphibian model in figure 8. Figure 11 shows the response to constant velocities of stretch with experimental data taken from the inferior oblique extraocular muscle in cats (ref. 17). The extraocular stretch receptors are functionally similar to spindles in their afferent nerve's response to varying degrees and rates of muscle stretch. They differ histologically, however, in that no encapsulated muscle fibers with nerve endings resembling muscle spindles with their nuclear bag region have been found (ref. 18). Instead, occasional myelinated fibers are seen making two or three spiral turns around a small muscle fiber. These organs, it is thought, act as the stretch sensor.

MULTIPLE INPUT-OUTPUT MAMMALIAN SPINDLE.—The values of the parameters for the multiple input-output spindle model with no motor stimulation were chosen as follows:

$$K=8 \text{ N/m}$$

$$B_1=B_2=0.12 \text{ N/m/sec}$$

$$B_3=B_4=0.06 \text{ N/m/sec}$$

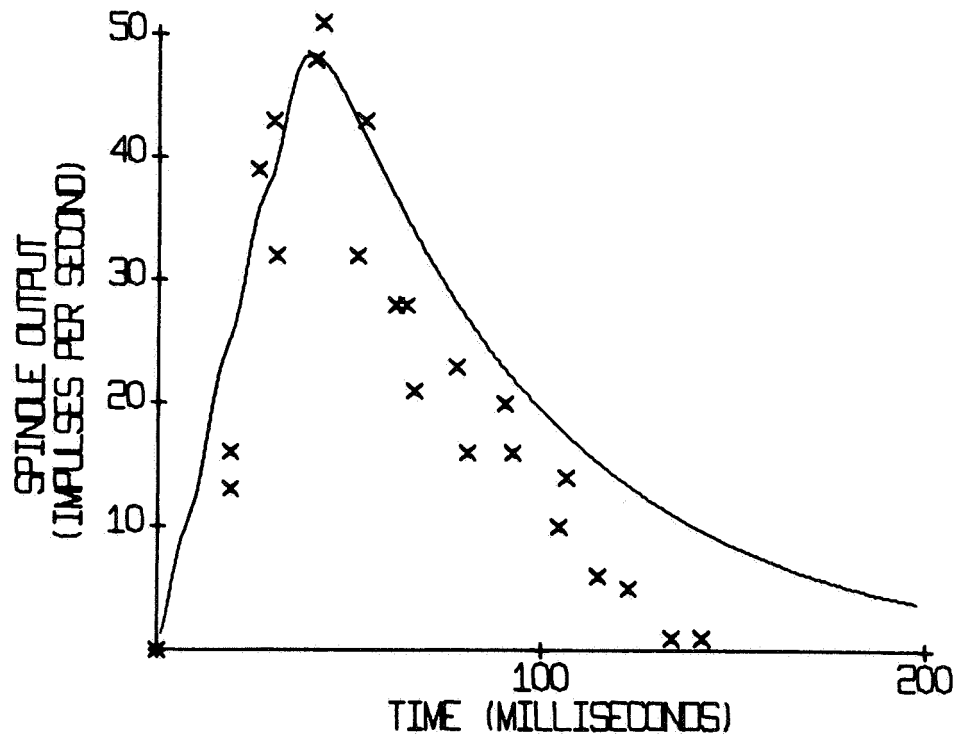


Figure 10.—Response of the mammalian spindle to four intrafusal muscle twitches. The continuous curve is the model response to the same input as in figure 5. The discrete points are redrawn from Eyzaguirre (ref. 16).

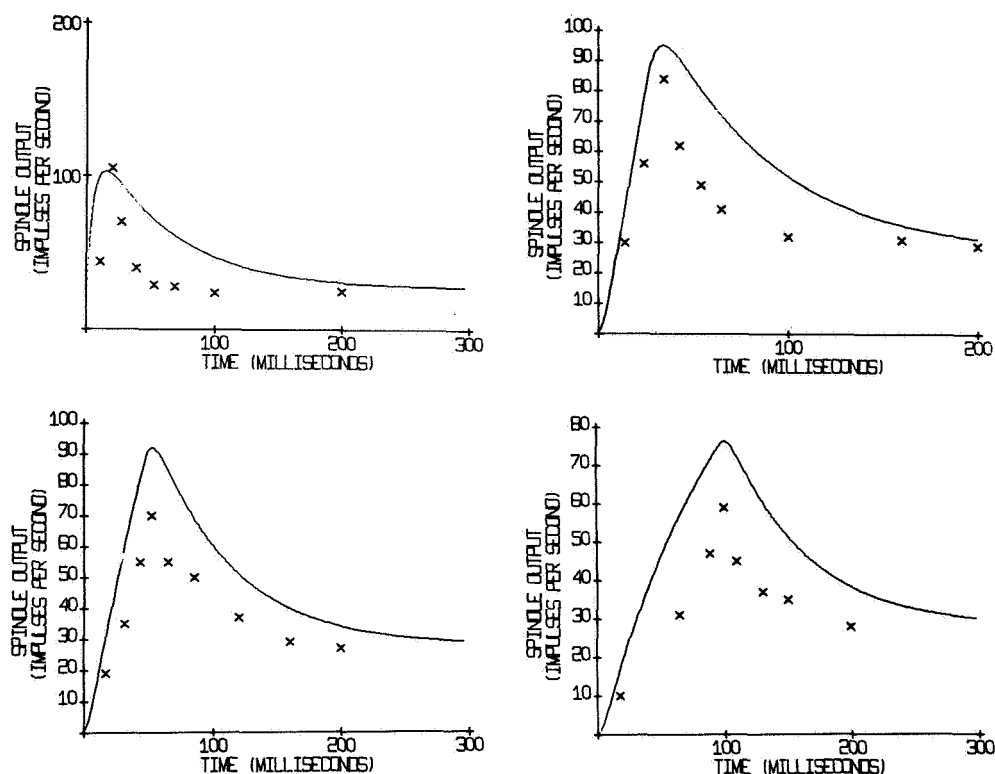


Figure 11.—Response of the mammalian spindle to constant velocity stretch. The continuous curves are the model response. The discrete points are redrawn from the experimental data of Bach-Y-Rita and Ito (ref. 17). The stretch velocities are (a) 200 mm/sec, (b) 100 mm/sec, (c) 60 mm/sec, and (d) 30 mm/sec.

We have assumed that the dynamic fiber stimulation increases the viscosity of the nuclear bag fibers by a factor of two, i. e., $B_1=B_2=0.24$ N/m/sec, whereas the static fiber stimulation increases the viscosity of the polar and presumably contractile regions of the chain fibers by a factor of two, i. e., $B_3=0.12$ N/m/sec. All the proportionality constants K_B , K_C , and K_{CS} were chosen equal to 100 pulses per sec/mm of stretch.

The results of the computer simulation of the multiple input-output spindle model for a constant velocity stretch input of 20 mm/sec for a duration of 100 msec are shown in figure 12. The output of the primary ending is shown in figure 12(a). This output is greatly enhanced by a simultaneous dynamic fiber stimulation (D-curve), whereas this output is essentially unaffected by a simultaneous static fiber stimulation (S-curve). The U-curve is the output of the nuclear bag branch of the primary ending with no simultaneous motor input.

The output of the nuclear chain branch of the primary afferent is shown in figure 12(b). This output, in our simulation, is also equal to the total output of the secondary endings. This response is most strongly increased by a simultaneous static fiber stimulation and less strongly so by a simultaneous dynamic fiber stimulation.

Another approach to the study of systems that are linear, or nearly so, is to apply sinusoidal stretch and measure the relative magnitude and phase of the response. Figure 13 shows the Bode plots for the amphibian model. Figure 14 shows those for the mammalian model as well as some data from cat's tibialis anterior (ref. 19). The lack of correlation between the phase characteristics of the model and the experimental data is probably largely

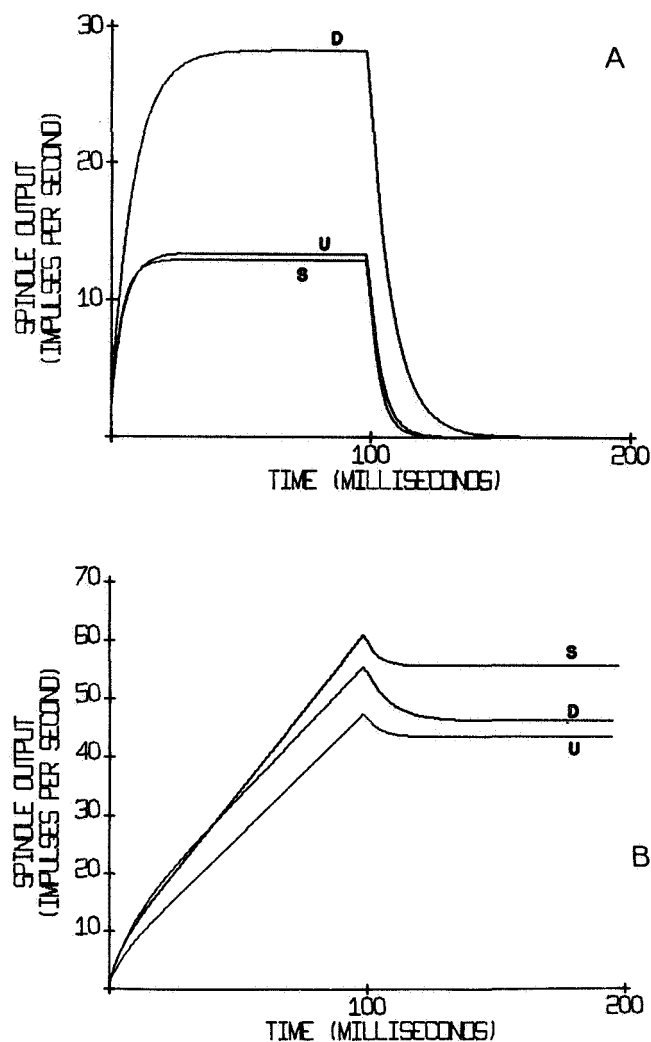


Figure 12. —Response of the multiple input-output mammalian spindle model to a constant velocity stretch of 20 mm/sec for a duration of 100 msec. U-unstimulated, D-simultaneous dynamic fiber stimulation, S-static fiber stimulation. (a) Output of the nuclear bag of the afferent ending. (b) Output of the nuclear chain part of the afferent endings.

attributable to the fact that the spindle was not sufficiently stretched to fire tonically throughout the cycle but instead behaved by "bursting" during brief intervals of each period. Although there were no data found in the literature for amphibian spindles, such data are available for crustacean stretch receptors and are shown in figure 15 (ref. 20).

DISCUSSION

AMPHIBIAN AND SINGLE INPUT-OUTPUT MAMMALIAN SPINDLE MODELS. — The data show that both models can provide a reasonably good fit to experimental results for appropriate parameter values. Although the parameters were chosen for these models from

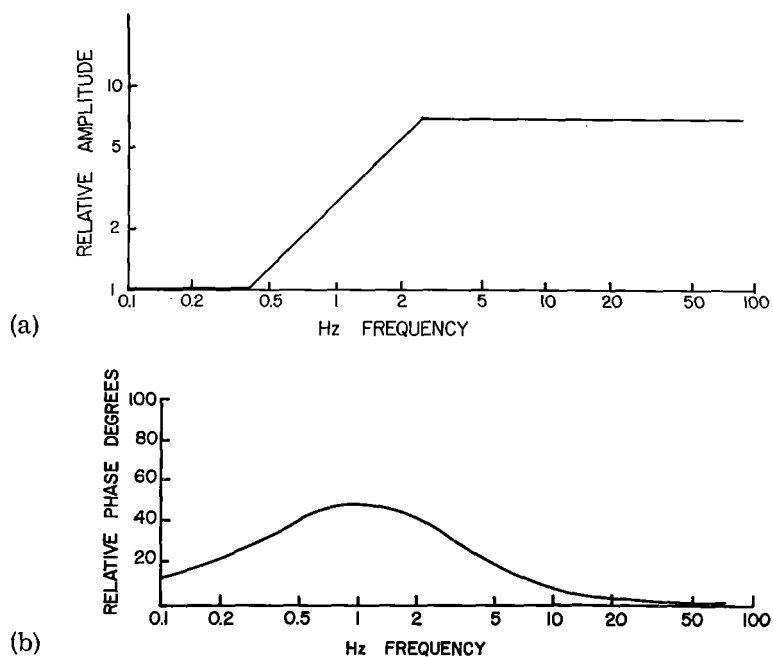


Figure 13.—Bode plots of the amphibian spindle model. (a) Relative amplitude, and (b) Relative phase.

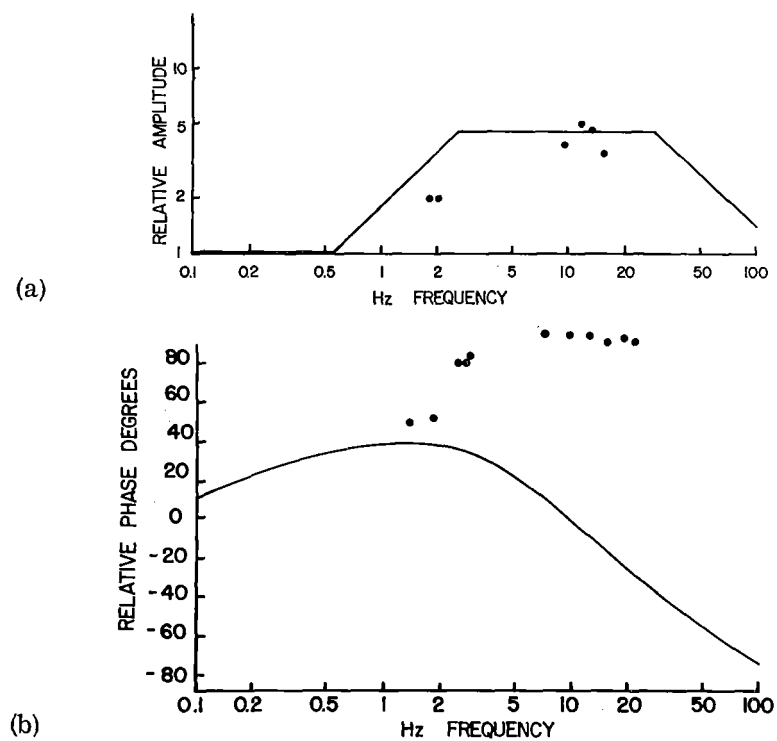


Figure 14.—Bode plots of the single input-output spindle model. (a) Relative amplitude. (b) Relative phase. The discrete points are the experimental values taken from Lippold and Redfearn (ref. 19).

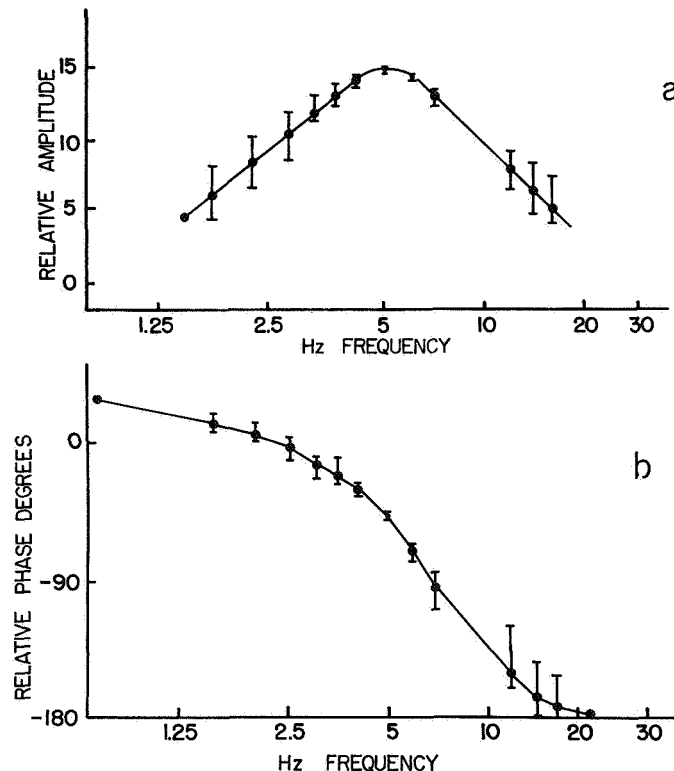


Figure 15.—Bode plots of the crustacean stretch receptor redrawn from Borsellino, Poppele, and Terzuolo (ref. 20). (a) Relative amplitude. (b) Relative phase.

one single experiment, the models qualitatively describe other experiments about as well as can be expected when one tries to fit several men with one pair of shoes.

However, certain discrepancies between model and experimental data of figures 7(a) and (b) and to a lesser extent (c) show a sudden decrease in the rate of spindle firing which is not explained by our model and cannot be easily incorporated. If this is truly a property of the receptor, then it is probably not a consequence of simple mechanical phenomena, but is more likely of some other origin; possibly something like stretch activation of intrafusal fibers which has been described as a property of some insect flight muscles (ref. 21).

A property not apparent in the experimental data presented here is the presence of an additional, longer time constant of the order of 2 seconds (refs. 16 and 22). Such behavior could be incorporated in a slightly more complex mechanical model, but this is not justified by the mechanical behavior of the physiological system. More likely this could result from fatigue and hyperpolarization at the level of the cell membrane (ref. 16).

A third divergent phenomenon is the pronounced upward concavity of the data in figure 11(d) during moderate rates of constant-velocity stretch. This property is not displayed by our model but could be incorporated by a number of nonlinear properties of which our springs, dashpots, and other linear assumptions are but approximations. A simple mechanism would be to have a decrease in mechanical resistance, either in elasticity or viscosity or both, at increasing lengths. Length-tension curves of inactive muscle show that this is clearly not the case for the elastic components. There is some evidence that viscosity may decrease with lengthening (ref. 23) and increase during active shortening (ref. 24), but it is probably not significant for our experiments here.

Two other possible loci for the nonlinearity are between receptor potential and applied stretch between receptor potential and firing rate; but here, too, the available data (refs. 14 and 15), which is quite limited and incomplete, indicate the absence of such a property or the presence of a nonlinear relationship with an effect inverse to that which we are seeking.

On the basis of the mammalian transfer function that has been presented, we can compute three of the five mechanical parameters if we know the other two. Such information would be quite interesting since it would allow an estimate of the force exerted by intrafusal contraction and the amount of extrafusal force required when the spindles are stretched passively. The former force is quite small and has only recently been measured by Diete-Spiff (ref. 25). His in vitro recordings of isolated cat's lumbrical muscle spindle contractions during direct electrical tetanic stimulations ranged from 0.21 to 4.8 milligrams at 20°C.

Unfortunately, there are no reported measurements of spindle mechanical parameters, but there are numerous measurements of muscle parameters; and these may be used as a starting point. The most common muscle model is essentially identical to the spindle model of figure 4. If we look at muscle more closely, we find that such a model is actually representative of a single sarcomere. Each muscle fiber is many sarcomeres linked in series, and each muscle is many fibers linked in series and parallel. In terms of gross muscle behavior, looking at the model in terms of many sarcomere elements is, by appropriate techniques, identical to the model of figure 4. The essential point brought out by considering individual sarcomere elements is that both the elasticity and viscosity of the total muscle will be proportional to the number of fibers in parallel, that is, to the muscle cross-sectional area, and inversely proportional to the number of elements in series, muscle length. We will therefore define, by equation (11), a geometric ratio G

$$G = \frac{XSA}{L} \quad (11)$$

where XSA is muscle cross-sectional area and L its length, which relates a muscle's structural parameters which characterize the muscle as a whole to its normalized parameters—the basic properties of the biological materials. This is illustrated by equation (12) where V_m is some mechanical parameter measured from an intact muscle and V_n is the normalized parameter characteristic of the muscle tissue.

$$V_m = GV_n \quad (12)$$

Table 1 summarizes some parameters estimated from the literature.

There is a considerable overlap in the ranges of the parameters among the preparations. The human parameters are lower but this is probably reasonable for an in vivo preparation.

There are excellent data available on spindle morphology (refs. 26 and 27). Using the data from Barker and Gidumal (ref. 26), we can compute a geometric ratio $G = 8.4 \times 10^{-6}$ for an average spindle in the cat's rectus femoris. Using this G and nominal values from table 1 we can use equation (12) to compute normalized mechanical parameters for the spindle. An example of such nominal values are shown in table 2.

In computing these numbers, the structural parameters K_T and K_F were chosen first and K_B , B_B , and B_F were then computed from the transfer functions, equations (3), (8), and (9). With the exception of the parallel elastic element K_F for which no values are included in table 1, the normalized parameters for the spindle fall within the range of normalized parameters for the intact muscle in table 1.

TABLE 1.—TYPICAL MUSCLE MECHANICAL PARAMETERS

Structural Parameters		Normalized Parameters		Geometric ratio G (Meters)	Preparation	Reference
Elasticity Newtons/meter	Viscosity Newtons/meter/sec	Elasticity Newtons/meter ²	Viscosity Newtons/meter ² /sec			
98-450	1.4-34	$0.83-3.8 \times 10^6$	$6.2-29 \times 10^4$	1.2×10^{-4}	frog sartorius	[25]
350-520	2.9-15	$2.9-4.3 \times 10^6$	$2.4-12 \times 10^4$	1.2×10^{-4}	frog sartorius	[22]
670-2000	8.1-67	$0.083-2.5 \times 10^6$	$.1-.83 \times 10^4$	8.1×10^{-3}	human arm	[26]
1900-5500	12-68	$2.8-8.2 \times 10^6$	$1.8-10 \times 10^4$	6.7×10^{-4}	rat anterior tibialis	[27]

TABLE 2.—NOMINAL SPINDLE
MECHANICAL PARAMETERS

Parameter	Structural Parameters	Normalized Parameters
K_T	10	1.19×10^6
K_F	0.45	5.35×10^4
K_B	4.0	4.76×10^5
B_B	0.12	1.44×10^4
B_F	0.13	1.51×10^4

We have assumed that the intrafusal series elasticity is relatively high and the parallel elasticity low. Our computations show us that the nuclear bag region must be considerably stiffer than the intrafusal muscle while their respective viscosities are similar.

With these parameters we can now estimate intrafusal force. The net elasticity of the three series springs is 0.625 N/m. Let us assume a spindle stretch of 0.35 millimeter. This can be justified in two ways. First, it is the value used by Shepherd and Ottoson (ref. 15). Second, a human gastrocnemius muscle, 30 centimeters long, is stretched 2 centimeters by dorsiflexion. If a spindle is 5 millimeters long and is stretched proportionately, we get a spindle stretch of about 0.33 millimeters.

With the above values, force required to stretch the spindle 0.35 millimeters is 2.2×10^{-4} newton which is on the order of 0.02 gram weight of force.² To impose a corresponding amount of stretch on the nuclear bag by intrafusal muscle contraction, the net intrafusal force must be the same. This agrees surprisingly well with the measurements of Diete-Spiff (ref. 25).

The actual force F_M generated in such a contraction would be 2.6×10^{-4} newton, the additional force being required to compress K_F . Furthermore, even in a slow stretch of 1 mm/sec, the viscous force of the intrafusal fiber is 2.1×10^{-4} newton, so that in both these cases, intrafusal elasticity contributes 20 percent or less to the total tension. This strengthens our case in neglecting its equivalent K_1 altogether in this amphibian model.

From the data presented, it is evident that either of the two models does an adequate job of describing the response to transient dynamic inputs. Figure 16 shows the amphibian model response to a step input. The instantaneous jump of the response in frequency is unrealistic (although it does not contradict the experimental data) and probably undesirable, but this sort of input is extremely unphysiological. Although the step response is a performance measure well beloved of the engineer, it should hold a somewhat less exalted position for the physiologist who might regard the ramp (constant velocity) input as a more realistic test of spindle system behavior.

²We can make a similar estimate of 0.08 grams force for a single skeletal muscle fiber (ref. 28).

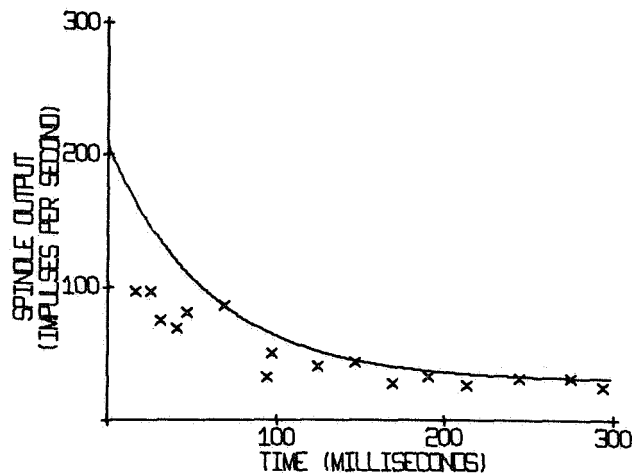


Figure 16. —The continuous curve shows the step response of the amphibian spindle model. The discrete points are taken from the experimental data of Lippold, Nicholls, and Redfearn (ref. 14) on the cat's muscle spindle.

The principal difference between the two models is, of course, their behavior at high frequencies, as clearly indicated in the Bode plots, figures 13 and 14. It is the high frequency falloff that makes the mammalian model to respond more respectably to a step than does the amphibian model. But, if, for a physiological range of transient inputs, the two models are equally good, as elements in the overall motor-control system, they may be significantly different. The mammalian model provides 90° of phase lag at high frequencies which might be important to system stability. However, even this seems doubtful. At 100 hertz, where the phase difference due to model dynamics is 90° , transport delays up the afferent and down the efferent nerves provide about 360° or more of phase lag in humans. Additional delays in the spinal cord and at the muscle would add to this. Furthermore, physiological tremor is in the range of 8-10 hertz (ref. 29), where the models are similar.

Thus, although the signal input-output mammalian model may be more accurate than the amphibian model as an isolated element, as an element in a more comprehensive system, the two models appear to be equivalent.

MULTIPLE INPUT-OUTPUT MAMMALIAN SPINDLE. —In order to clarify our terminology, yet remain reasonably consistent with the physiological literature, let us define three terms, dynamic sensitivity, static sensitivity, and bias. The first two refer to the steady-state behavior after the transients have died out. The dynamic sensitivity is a measure of that component of the spindle afferent output which is proportional to spindle stretch velocity. It has units of pps/mm/sec. The static sensitivity is a measure of that component of the spindle afferent output which is proportional to spindle stretch. It has units of pps/mm. The bias is the tonic component of the spindle output that is independent of stretch. It has units of pps. With reference to these three quantities, the experimental results of Matthews and his associates, Bessou and Laporte, and others (refs. 8, 30, 31, and 32) may be briefly summarized as shown in table 3.

With reference to figure 12, the dynamic and static sensitivities are to be interpreted as follows. Suppose an unstimulated spindle at a length just below its threshold of firing is suddenly stretched at a constant velocity for 100 msec and then held at this new length. If its response looks like one of the curves of figure 12(a), the spindle has a dynamic sensitivity defined by the height of the firing plateau divided by the velocity of stretch. If its response

TABLE 3.—MULTIPLE INPUT-OUTPUT MAMMALIAN SPINDLE

		Unstimulated	Dynamic Fiber Stimulation	Static Fiber Stimulation
Primary Endings	Bias	Variable	Increase	Greater Increase
	Dynamic Sensitivity	Significant	Increase	No Change or Slight Decrease
	Static Sensitivity	Significant	Increase	Greater Increase
Secondary Endings	Bias	Variable	No Change	Increase
	Dynamic Sensitivity	Slight	No Change	No Change
	Static Sensitivity	Significant	No Change	Increase

looks like one of the curves of figure 12(b), the spindle has a static sensitivity defined as the spindle output after the stretch has ended, divided by the change in the spindle length. Note that this ratio is very close to the ratio of spindle output to change in spindle length, at any time, even during a constant velocity stretch.

Above we assumed that spindle was silent prior to the stretch just below its firing threshold. If any of the fusimotor nerves were stimulated, the spindle might begin to fire. This would represent a change in bias.

For our discussion of the experimental data taken from literature and the results of our model simulation, we will take the unstimulated primary and secondary outputs as our reference. In table 3, note that the primary endings have both significant dynamic and static components, while the secondary endings have very small dynamic but significant static components.

The effects of γ fiber stimulation on the spindle response are essentially bimodal with responses falling into one of the two categories. Those fibers that have the most pronounced effects on the dynamic sensitivity of the spindle afferents are called dynamic fusimotor fibers while those that have strong effect on the static output are called the static fusimotor fibers.

From table 3 we can see that the dynamic fiber stimulation increases bias, dynamic sensitivity, and static sensitivity of the primary endings but has negligible effect on the secondary endings. By contrast, the static fiber stimulation has strong influence on bias and static sensitivity of both primary and secondary endings but has little effect on dynamic sensitivity of either afferents.

On the basis of their observations, Matthews and Laporte and their associates (refs. 8 and 33) have suggested that static fibers correspond to fibers which innervate intrafusal muscle chain fibers (Boyd's γ_2 fibers) and dynamic fibers to fibers which innervate intrafusal muscle bag fibers (Boyd's γ_1 fibers). The γ_2 fibers affect both primary and secondary endings since both of these endings synapse on the chain fibers. The γ_1 fibers influence only the primary endings since only primary endings synapse on the bag fibers. Furthermore, we

would expect the greatest dynamic sensitivity from the nuclear bag portion of the bag fibers where the viscosity is thought to be relatively low and the elasticity high.

Figure 12 shows the results of our computer simulation for a constant velocity stretch input of 20 mm/sec for a duration of 100 msec without and with simultaneous γ -motor-fiber stimulations. Because we did not include any force generators in the model, we do not get any change in bias during stimulation since this is caused by the actual contraction of the intrafusal fibers. The changes in sensitivity, however, are caused not by the contraction of the intrafusal fibers, which is just a linearly additive component of the response, but by changes in the physical parameters. For the purpose of this simulation, we have assumed that the fusimotor stimulation increases the viscosity of the contractile regions of the bag and chain fibers by a factor of two.

The parameter values used in this simulation are close to those used in the earlier simulations of the amphibian and single input-output mammalian spindle models. The time constant was shortened slightly in order to shorten the transient phase of the response and to make the dynamic and static properties more evident.

The only discrepancy between our model output and the experimental observations is a slight increase in the static sensitivity of the secondary endings due to dynamic fiber stimulation (figure 12(b) curve D and table 3). This error will be corrected in a more comprehensive model which will include the force generators and the elastic components.

SUMMARY

- (1) Three lumped parameter models of muscle stretch receptors have been described.
- (2) It is shown that simple, linear, mechanical models account for many of the behavior characteristics exhibited by muscle spindles but are inadequate for a complete model. These inadequacies are discussed.
- (3) The mechanical parameters are discussed, and it is shown that the passive values used in the model are reasonable relative to values measured from physiological preparations.
- (4) For the mechanical parameters suggested, it is shown that the necessary forces of intrafusal contractions are extremely low.
- (5) The behavior of the amphibian and single input-output mammalian models is compared with data from mammalian, amphibian, and crustacean preparations. These two models are shown to be physiologically equivalent from a systems point of view.
- (6) A simple multiple input-output model of mammalian spindle is shown to predict qualitatively most of the significant characteristic responses due to unstimulated stretch input and during static and dynamic fusimotor fiber stimulations.

APPENDIX A—INPUT-OUTPUT EQUATION FOR AMPHIBIAN SPINDLE

In the lumped mechanical model of the amphibian spindle, figure 2, define

$$X_R \triangleq X_M - X_D \quad (A-1)$$

where both X_D and X_M are measured with reference to their position when $F=0$. The equations of motion are

$$F = K_2 (X_M - X_D) \quad (A-2)$$

$$F = K_1 X_D + B s X_D + F_M \quad (A-3)$$

Eliminating F from these equations and solving for X_D , we obtain

$$X_D = \frac{A_2}{T_o s + 1} X_M - \frac{A_o}{T_o s + 1} F_M \quad (A-4)$$

where

$$A_o = \frac{1}{K_1 + K_2}$$

$$A_2 = \frac{K_2}{K_1 + K_2}$$

$$T_o = \frac{B}{K_1 + K_2}$$

Substituting (A-4) in (A-1), we obtain

$$X_R = A_1 \frac{T_1 s + 1}{T_o s + 1} X_M + \frac{A_o}{T_o s + 1} F_M \quad (A-5)$$

where

$$A_1 = \frac{K_1}{K_1 + K_2}$$

$$T_1 = \frac{B}{K_1}$$

In terms of α , as defined in equation (2), equation (1) may be rewritten as

$$f = \frac{k_f N_D}{X_{Do}} [X_D + \alpha X_R] \quad (A-6)$$

By substituting (A-4) and (A-5) into (A-6), we obtain

$$f = K_X \frac{T_2 s + 1}{T_o s + 1} X_M + \frac{K_F}{T_o s + 1} F_M \quad (A-7)$$

where

$$K_X = \frac{k_f N_D}{X_{Do}} \left[\frac{\alpha K_1 + K_2}{K_1 + K_2} \right]$$

$$K_F = \frac{k_f N_D}{X_{Do}} \left[\frac{\alpha - 1}{K_1 + K_2} \right]$$

$$T_2 = \frac{\alpha B}{K_1 + K_2}$$

As noted in the text, our simulations are done with $K_1=0$. Notice for $K_1=0$, the form of the equation (A-7) is unchanged. The new parameters are

$$K_X = k_f \frac{N_D}{X_{Do}}$$

$$K_F = k_f \frac{N_D}{X_{Do}} \frac{\alpha - 1}{K_2}$$

$$T_2 = \alpha \frac{B}{K_2}$$

$$T_o = \frac{B}{K_2}$$

With these substitutions in equation (A-7), the location of the pole, relative to the zero, in the first transfer function on the right-hand side of (A-7), is a function of α alone and is independent of the mechanical parameters.

APPENDIX B—INPUT-OUTPUT EQUATION FOR MAMMALIAN SPINDLE

In the lumped mechanical model of the mammalian spindle (fig. 3) X_1 , X_2 , and X_M are measured from the rest condition when $F=0$ and $F_M=0$. Since this is a linear system and superposition applies, we can obtain the transfer function by considering one source at a given time.

First assume $F_M=0$, then the equations of motion are

$$F = K_T X_1 \quad (B-1)$$

$$F = (K_F + B_F s) (X_2 - X_1) \quad (B-2)$$

$$F = (K_B + B_B s) (X_M - X_2) \quad (B-3)$$

Eliminating F and X_1 in these equations, we get

$$\frac{K_B + B_B s}{K_T} (X_M - X_2) = \frac{K_F + B_F s}{K_F + K_T + B_F s} X_2 \quad (B-4)$$

Rearranging,

$$\frac{X_2}{X_M} = \frac{(K_B + B_B s) (K_F + K_T + B_F s)}{K_T (K_F + B_F s) + (X_B + B_B s) (K_F + K_T + B_F s)} \quad (B-5)$$

Since the firing rate in the afferent nerve is assumed to be proportional to $X_M - X_2$, define the transfer function

$$T_X(s) = \frac{X_M - X_2}{X_M} \quad (B-6)$$

From (B-5) and (B-6) we get

$$T_X(s) = \frac{K_T}{K_B} \frac{s + K_F/B_F}{s^2 + (\frac{1}{T_1} + \frac{1}{T_2}) s + \frac{1}{T_1 T_2}} \quad (B-7)$$

where

$$\frac{1}{T_1} + \frac{1}{T_2} = \frac{K_T B_F + K_B B_F + K_F B_B + K_T B_B}{B_B B_F}$$

$$\frac{1}{T_1 T_2} = \frac{K_T K_F + K_B K_F + K_B K_T}{B_B B_F}$$

Now assume $X_M=0$ and $F_M \neq 0$. The equations of motion are

$$F = K_T X_1 \quad (B-8)$$

$$F = (X_2 - X_1)(B_F s + K_F) + F_M \quad (B-9)$$

$$F = -X_2 (B_B s + K_B) \quad (B-10)$$

Eliminating F and X_1 in these equations, we get

$$\begin{aligned} \frac{-X_2}{F_M} &= \frac{K_T / B_B B_F}{s^2 + \left(\frac{1}{T_1} + \frac{1}{T_2}\right)s + \frac{1}{T_1 T_2}} \\ &= T_F(s) \end{aligned} \quad (B-11)$$

The total transfer function is obtained by combining (B-7) and (B-11)

$$X_M - X_2 = T_X(s) X_M + T_F(s) F_M \quad (B-12)$$

APPENDIX C—MULTIPLE INPUT-OUTPUT MAMMALIAN SPINDLE

From figure 6, the equations of motion are

$$B_1 s (X_M - X_A) = B_2 s (X_A - X_B) + B_3 s (X_A - X_C) \quad (C-1)$$

$$B_2 s (X_A - X_B) = K X_B \quad (C-2)$$

$$B_3 s (X_A - X_C) = B_4 s X_C \quad (C-3)$$

From (C-2) and (C-3)

$$X_B = \frac{B_2 s}{B_2 s + K} X_A \quad (C-4)$$

$$X_A = \left(1 + \frac{B_4}{B_3}\right) X_C \quad (C-5)$$

From (C-1), (C-4), and (C-5), we can solve for X_C in terms of X_M

$$X_C = G_C \frac{s + \frac{1}{T_1}}{s + \frac{1}{T_0}} X_M \stackrel{\Delta}{=} T_C(s) X_M \quad (C-6)$$

where

$$G_C = \frac{B_1 B_3}{B_4 (B_2 + B_3)}$$

$$\frac{1}{T_o} = \frac{(B_1 + B_2) (B_3 + B_4) + B_3 B_4}{(B_1 + B_3) B_2 B_4} K$$

$$\frac{1}{T_1} = \frac{K}{B_2}$$

From (C-4), (C-5), and (C-6) we get

$$X_B = G_B \frac{s}{s + \frac{1}{T_o}} X_M \stackrel{\Delta}{=} T_B(s) X_M \quad (C-7)$$

where

$$G_B = \frac{B_1 (B_3 + B_4)}{B_4 (B_2 + B_3)}$$

The output of the primary ending f_p is proportional to the stretch of both the bag and chain fibers. From (C-6) and (C-7), we get

$$f_p = [K_B T_B(s) + K_C T_C(s)] X_M \quad (C-8)$$

where K_B and K_C are constants of proportionality. The output of the secondary ending f_s results from the stretching of the chain fibers only.

From (C-6) we get

$$f_s = K_{CS} T_C(s) X_M \quad (C-9)$$

where K_{CS} is a constant of proportionality.

REFERENCES

1. Houk, J.; Cornew, R.; and Stark, L.: A Model of Adaptation in Amphibian Spindle Receptors. *J. Theoret. Biol.*, vol. 12, 1966, pp. 196-215.
2. Barker, D.: The Innervation of the Muscle Spindles. *Quart. J. Micr. Sci.*, vol. 89, 1948, pp. 143-186.
3. Eyzaguirre, C.: The Electrical Activity of Mammalian Intrafusal Fibers. *J. Physiol.*, vol. 150, 1960, pp. 169-185.
4. Katz, B: The Terminations of the Afferent Nerve Fiber in the Muscle Spindle of the Cat. *Phil. Trans. Royal Soc.*, London, 243B, 1961, pp. 221-240.
5. Stark, L.: Neurological Feedback Control Systems. Ch. 4 in *Advances in Bioengineering and Instrumentation*. F. Alt, ed., Plenum Press, N. Y., 1966.
6. Boyd, I. A.: The Motor Innervation of Mammalian Muscle Spindles. *J. Physiol.*, vol. 159, 1961, pp. 7-9.
7. Barker, D., ed.: *Symposium on Muscle Receptors*. Hong Kong University Press, 1962.
8. Granit, R., ed.: *Nobel Symposium Muscular Afferents and Motor Control*. John Wiley & Sons, Inc., 1966.

9. Matthews, P. B. C.: Muscle Spindles and Their Motor Control. *Physiol. Rev.*, vol. 42, 1964, pp. 219-288.
10. Ruffini, A.: On the Minute Anatomy of the Neuro-Muscular Spindles of Cat, and on their Physiological Significance. *J. Physiol.*, vol. 23, 1898, pp. 190-208.
11. Boyd, I. A.: The Nuclear-bag Fiber and Nuclear-chain Fiber Systems in the Muscle Spindles of the Cat. Symposium on Muscle Receptors, D. Barker, ed., Hong Kong University Press, 1962, pp. 185-190.
12. Andersson, B.; Lennerstrand, G.; and Thoden, U.: Cat Muscle Spindle Model. Digest of the 7th International Conference on Medical and Biological Engineering, Stockholm, 1967, p. 290.
13. Kuo, B.: Automatic Control Systems. Prentice-Hall, N. J., 1962.
14. Lippold, O.; Nicholls, J.; and Redfearn, J.: Electrical and Mechanical Factors in the Adaptation of A Mammalian Muscle Spindle. *J. Physiol.*, vol. 153, 1960, pp. 209-217.
15. Shepherd, G.; and Ottoson, D.: Response of the Isolated Muscle Spindle to Different Rates of Stretching. *Symp. Quant. Biol.*, 30, Sensory Receptors, Cold Springs Harbor Lab. of Quant. Biol. 1965.
16. Eyzaguirre, C.: The Motor Regulation of Mammalian Spindle Discharges. *J. Physiol.* vol. 150, 1960, pp. 186-200.
17. Bach-y-Rita, P.; and Ito, F.: Properties of Stretch Receptors in Cat Extraocular Muscles. *J. Physiol.*, vol. 186, 1966, pp. 663-688.
18. Cooper, S.; and Fillenz, M.: Afferent Discharges in Response to Stretch From the Extra-Ocular Muscles of the Cat and Monkey and the Intervation of these Muscles. *J. Physiol.*, vol. 127, 1955, pp. 400-413.
19. Lippold, O.; and Redfearn, J.: The Effect of Sinusoidal Stretching upon the Activity of Stretch Receptors in Voluntary Muscle and Their Reflex Response. *J. Physiol.*, vol. 144, 1958, pp. 323-386.
20. Borsellino, A.; Poppele, R.; and Terzuolo, C.: Transfer Functions of the Slowly Adapting Stretch Receptor Organ of Crustacea. *Symp. Quant. Biol.*, 30, Sensory Receptors. Cold Springs Harbor Lab. Quant. Biol., 1965.
21. Pringle, J.: The Excitation and Contraction of the Flight Muscles of Insects. *J. Physiol.*, vol. 108, 1949, pp. 226-232.
22. Ottoson, D.; and Shepherd, G.: Receptor Potentials and Inpulse Generation in the Isolated Spindle During Controlled Extension. *Symp. Quant. Biol.*, 30, Sensory Receptors. Cold Springs Harbor Lab. Quant. Biol., 1965.
23. Jewell, B.; and Wilkie, D.: An Analysis of the Mechanical Components in Frogs Striated Muscle. *J. Physiol.*, vol. 143, 1958, pp. 515-540.
24. Buchthal, F.; and Rosenfalck, C.: Dynamic Elasticity in the Initial Phase of an Isotonic Twitch. *Acta Physiol. Scand.*, vol. 149, 1960, pp. 198-210.
25. Diete-Spiff, K.: Tension Development by Isolated Muscle Spindles of the Cat. *J. Physiol.*, vol. 193, 1967, pp. 31-43.
26. Barker, D.; and Gidumal, J.: The Morphology of Intrafusar Muscle Fibers in the Cat. *J. Physiol.*, vol. 157, 1961, pp. 513-528.
27. Eldred, E.; Bridgman, C.; Swett, J.; and Eldred, B.: Quantitative Comparisons of Muscle Receptors of the Cat's Medial Gastrocnemius Soleus and Extensor Digitorum Brevis Muscles. Symposium on Muscle Receptors. D. Barker, ed., Hong Kong Univ. Press, 1962.

28. Woodbury, J.; Gordon, A.; and Conrad, J.: Muscle. Ch. 5 in Physiology and Biophysics, T. Ruch, and H. Patton, ed., W. B. Saunders Co., Philadelphia, 1965.
29. Lippold, O. C. J.; Redfearn, J. W. T.; and Vuco, J.: The Rhythmical Activity of Groups of Motor Units in the Voluntary Contraction of Muscle. *J. Physiol.*, vol. 137, 1957, pp. 473-487.
30. Crowe, A.; and Matthews, P. B. C.: The Effects of Stimulation of Static and Dynamic Fusimotor Fibers on the Response to Stretching of the Primary Endings of Muscle Spindles. *J. Physiol.*, 174, 1964, pp. 109-131.
31. Appelberg, B.; Bessou, P.; and Laporte, Y.: Action of Static and Dynamic Fusimotor Fibers on Secondary Endings of Cat's Spindles. *J. Physiol.*, vol. 185, 1966, pp. 160-171.
32. Brown, M. C.; Engberg, I.; and Matthews, P. B. C.: Fusimotor Stimulation and the Dynamic Sensitivity of the Secondary Ending of the Muscle Spindle. *J. Physiol.*, vol. 189, 1967, pp. 545-550.
33. Jansen, J. K. S.; and Matthews, P. B. C.: The Central Control of the Dynamic Response of Muscle Spindle Receptors. *J. Physiol.*, vol. 161, 1962, pp. 357-378.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D. C. 20546
OFFICIAL BUSINESS

FIRST CLASS MAIL



POSTAGE AND FEES PAID
NATIONAL AERONAUTICS
AND SPACE ADMINISTRATION

POSTMASTER: If Undeliverable (Section 1103, Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

— NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546